Universitá degli Studi di Napoli
Federico II
Facoltá di Scienze MM. FF. NN.
Corso di Laurea in Fisica

The Multiplicity Function from Photometric Redshifts

Tesi di Laurea

Relatori:

Ch.mo Prof. Giuseppe Longo
Dott. Elisabetta De Filippis
Dott. Maurizio Paolillo

Candidato:

Magda Guglielmo
matricola: 60/928

Anno Accademico 2008/2009

Contents

List of Figures

List of Tables

Introduction

A galaxy is the environment in which stars are born and die, and distant galaxies are beacon that enable us to probe the distant universe. Our galaxy, the MilkyWay, is only one out of billions of such systems in the observable universe; furthermore, it is just a member of a group, the Local Group. Therefore, galaxies structures are an important field of study in cosmology; it has in fact been observed that in the Universe, objects tend to do not be isolated: stars form stellar clusters and then form galaxies. Galaxies themselves make groups, groups and clusters form superclusters

Once again the Local Group is a good example to understand such phenomenon, well described by Hubble, at the time of the first distance determinations and redshift measurements. The mechanism which controlled galaxy formation represents a central theme in modern cosmology. Among the main problems which have get to be solved: galaxies were formed before or after the large scale cosmic structures?

In other words, did massive clouds fragment in smaller clouds from which formed or, rather, galaxies formed first only of the mutual gravitational interaction made them group together to form structures? However, it is broadly recognized that studying galaxies clustering brings forth important clues of the conditions at the age of recombination.

It is for the mentioned reasons that the quest for the "perfect" galaxy group finding algorithm, started several years ago, is still an important open field of research.

Such "perfect finding algorithm", still does not exists, however, its specification and expectation are well known, it is in fact desirable that such algorithm results to be easy to apply, capable to handle as many selection effects as possible, yield reproducible results, without building in strong preconceptions about the group dynamics. Ideally it should be possible to vary

all selection criteria according to the specific needs.
It is here presented a new modified version of the friends-of-friends (FOF) structure finding algorithm (Huchra & Geller 1982), designed specifically to locate groups or clusters of galaxies in photometric redshift datasets. The main objective of this thesis is to show that this "new friends-of-friends" (hereafter New-FOF) algorithm yields results almost identical to the original FOF, if applied to a spectroscopic redshift dataset, and a rather conservative catalogue of structures, in case of a dataset with simulated photometric redshifts, when their errors are taken into account.
In order to achieve this purpose, after the mandatory introduction on galaxies groups morphology and origin, the basis of the original FOF algorithm are presented, together with the reasons that brought us and many other around the globe to extend such finding algorithm to photometric redshifts. The algorithm performance (catalogue size and composition) are then compared against with the performance of the standard FOF algorithm, firstly on a simulated dataset photometric redshifts, created using the 3D simulator developed for the GODFinGER (Galaxy Objects Detection Finalized to Groups Extensive Recognition) project.
Secondly, following the same methodology of parallel performance assessments of the New FOF against the standard FOF; catalogue obtained by observed data are then compared in size and composition. The New FOF algorithm (source code is attached in appendix.), it is realized in standard ANSI C to be platform independent, however, in this thesis it was compiled under a LINUX system.

Chapter 1

Formation of Structure in the Universe

Nowadays, it is well known that every cosmological model, in order to describe our universe has to comply with so called *Cosmological Principle* which states that the universe is homogeneous and isotropic on the large scale ($\geq 100\ Mpc$). This assumption (as it will be clarified later), is based on the well known property of translations and rotations invariance. In fact, to consider the universe homogeneous is equivalent to consider the galaxies and, more in general the mass, equally distributed in universe, and therefore an operation of translation from a galaxy to another would leave constant the galactic distribution taken into account. Moreover, the isotropic property of the universe represents the rotation invariance around one observation point. In few words, it does not exists any privileged direction in the space. On the other hand, when considering scales smaller than 100 Mpc, galaxies are not homogeneously distributed, rather they present themselves grouped in gravitational structures. Observing that all models willing to describe the origin of the universe must reproduce the "skeleton of the cosmic structures", it is clear that the understanding of galaxy distribution is among the fundamental purposes of cosmology.

1.1 Elements of Cosmology

Before proceeding with this exposition, it is important to shortly describe some of the main cosmological parameters.

1.1.1 Hubble's law

In the 1929's at the *National Academy of Sciences in Washington*, Edwin Hubble presented observational evidence that the spectrum of the galaxies was shifted to the red band. His experimental evidence was based on distance measurements obtained through Cepheids[1] for an handful of galaxies. Indicating with λ_e the wavelength of the radiation emitted by a source and with λ_o the observed wavelength, the quantity

$$\delta\lambda = \lambda_o - \lambda_e \tag{1.1}$$

is the so called Doppler shift caused by the fact that the galaxies are moving at a velocity expressed by the relationship

$$v = c\frac{\delta\lambda}{\lambda_e} \tag{1.2}$$

where c represents the speed of light. Hubble noted that there is a proportional relationship between galaxies distance and motion speed:

$$V = cz = H_0 D \tag{1.3}$$

where $z = \frac{\delta\lambda}{\lambda}$ represents the redshift of the spectral lines of the galaxy taken into account. H_0 represent a proportion constant with dimensions $kms^{-1}Mpc^{-1}$. It is worth to underline that the redshift it is not just due to a simple motion of the galaxies, but it is linked to a phenomenon known as "Hubble flow".
Such phenomenon finds explanation into the global recession speed of the Universe due to its expansion. Moreover, given the Hubble law unwavering with the space, it is possible to affirm that while in the universe expand the relative distances ratio is a constant, or in other words that the universe is always the same, the only thing varying with time being its scale. The above relationship can be also used to infer the distance of a galaxy on the

[1] The relationship between Cepheids ' period P, and its mean absolute magnitude M_v has been empirically derived. The period-luminosity relationship was:

$$M_v = -2.81\log(P) - (1.43 \pm 0.1)$$

with P measured in day.

basis of its radial speed:

$$D_z = H_0^{-1} cz \tag{1.4}$$

where with D_z is indicated the distance derived from the *redshift* z.

1.1.2 Scale Factors

In first instance it is possible to consider the galaxies in the universe as homogeneously and isotropically distributed and interacting through their gravitational field. According to the mentioned hypothesis, defined as t_0 and l_0 the initial conditions (present time and distance), it is possible the define the *scale factor*, as the quantity:

$$a(t) = \frac{l(t)}{l_0} \tag{1.5}$$

where $l(t)$ is the distance at the time t. It is also possible to normalize all the distances posing $a(t_0) = 1$. Furthermore, given the above hypothesis it is also possible to write the total mass density ρ conservation equation as:

$$\frac{d\rho}{dt} + 3\frac{\dot{a}}{a}\rho = 0 \tag{1.6}$$

From eq. 1.6 it is also possible to yield the Einstein equation (or the Newton equivalent) in terms of ρ namely as:

$$\frac{1}{a}\frac{d^2a}{dt^2} = -\frac{4\pi G}{3}\rho. \tag{1.7}$$

Making use of the above defined parameter $a(t)$ it is also possible to write the eq. 1.3 in another form:

$$H = \frac{\dot{a}}{a} = \frac{\dot{l}}{l} \tag{1.8}$$

this equation provides an experimental measurement of the ratio H. So, we can express the Hubble constant H_0 as:

$$H_0 = \frac{\dot{l_0}}{l_0} = \frac{\dot{a_0}}{a_0} \tag{1.9}$$

Moreover, given the last expression, it is possible to write a more general

definition of the *redshift* for a galaxy placed at the distance l

$$1 + z = \frac{1}{a} \tag{1.10}$$

1.1.3 The Hubble's constant

As mentioned above (Eq. 1.3), H_0 represents the proportionality constant between the velocity and the distance of a galaxy; this equation allows to estimate the present value of the Universe expansion. In paragraph 1.1.1 it already mentioned that the Hubble's constant has the physical dimension of $[t^{-1}]$; taking into account this observation it is possible to derive the so called *Hubble's time* (t_{Hubble}) as:

$$t_{Hubble} = \frac{1}{H_0}. \tag{1.11}$$

As it is possible to observe from the relationship 1.11, the Hubble time is the inverse of the Hubble constant, and it represents the time the universe needed to reach its actual configuration, hence, it is possible to say that t_{Hubble} represents the Universe's age. Therefore, t_{Hubble} represents the upper bound limitation for the age of all the celestial objects in the space.
Given the last, argument it is needless to say that the evaluation the exact value of such constant is an important field of research and still an open question not easy to solve.
In fact, determining the exact value of H_0, requires the availability of a method to gather the distance of galaxies independently from their redshift, a datum which is available only for few objects; to obtain such information for a large number of galaxies is not trivial.
Galaxies are far away objects and hence, it is not possible to calculate the distance parallax[2], and it is necessary to use methods based on the distance indicator objects, so called standard candles, i.e, which objects are characterized by an intrinsically known brightness.
One example of such kind of objects is the variable stars called Cepheids. They are good distance indicators because they exhibit an evident relationship between their period of variability, and their luminosity.

[2]By definition, an object at a distant of one parsec has a parallax equal to one arcsec. The object, in this case a galaxy, at a distant of one Megaparsec has a parallax too little to be measurable.

The standard candles objects are useful when all other geometric based methods such as the parallax, fail. It is necessary to stress that in order to fulfill the Hubble's law, distances must be large enough, cosmic recession speeds should be larger than the radial components of the peculiar ones, namely, larger than the peculiar velocities of the galaxies, due to the interaction with other galaxies. The average value of velocity is

$$v \sim 10^3 km\ s^{-1}$$

which correspond to a distant $D \geq 10^3 H_0^{-1} Mpc$.
A consequence of this rather complex procedure is that the value of this constant is still affect by considerable uncertainty. Just as a curiosity it is mentioned that since the first measurement of H_0, provided by Hubble himself ($H_0 \approx 500\ kms^{-1}Mpc^{-1}$); this value has been slowly reduced and now it falls in the range 50 e 100 $km\ s^{-1}Mpc^{-1}$. The most recent measures (performed using Hubble Space Telescope) estimated such value as: $H_0 = 71 \pm 4\ km\ s^{-1}\ Mpc^{-1}$.
In order to deal with such uncertainty a *Hubble parameter h*, defined as

$$H_0 = 100\ h\ km\ s^{-1}Mpc^{-1}.$$

1.1.4 Density Parameter

Considering the Universe to be a homogeneous and isotropic distribution of particles (in this case, *galaxies*), it is possible to define the critical density parameter, ρ_c. By definition, ρ_c is the density of an uniform self-gravitating sphere isotropically expanding at rate H with equal kinetic and gravitational potential energy.

$$\rho_c = \frac{3H^2}{8\pi G} \tag{1.12}$$

please observe that since H is a function of time, then so is ρ_c.
The value of this parameter is very important; in fact it determines the limit value to the expansion of the Universe: a universe, whose density is $\rho_c(t)$, stops expanding and contract down to a future singularity. On the other hand a density value less than $\rho_c(t)$, yields to an universe that expands forever.
Therefore, it is convenient to express the density of the Universe by introducing the *density parameter* Ω, defined as the ratio between the total density

and ρ_c:

$$\Omega(t) = \frac{\rho}{\rho_c} \tag{1.13}$$

Ω includes the contribution of several different kinds of matter: baryonic components, dark matter, dark energy, etc. The baryonic density, expressed as Ω_b, is the normal matter (protons, electrons neutrons, etc.). The dark matter is an unknown and not yet well understood form of matter. It is broadly accepted that it is not baryonic matter and that it interacts with the normal matter through the gravitational interaction. Dark matter, even though it is not directly observable, it is detectable only through its gravitational effect. The dark energy, Ω_L is a hypothetical form of energy that permeates all the space and tends to increase the expansion rate of the Universe, hence it is characterized by a negative pressure.
The knowledge of these terms allows us to determine the shape of the universe:

$\Omega_0 = 1 \qquad \rho = \rho_c \quad$ Universe is flat and infinitely extended

$\Omega_0 > 1 \qquad \rho > \rho_c \quad$ Universe is closed and finite

$\Omega_0 < 1 \qquad \rho < \rho_c \quad$ Universe open and infinitely extended

1.1.5 The Cosmological Constant

The cosmological term was introduced by Einstein when he applied general relativity for the first time to cosmology. In his 1917 paper, he found the first cosmological solution for a consistent theory of gravity. In spite of its drawbacks, this bold step can be regarded as the beginning of modern cosmology. Due to his belief in favor of static Universe, in order to find a static solution for his equation, Einstein introduced therefore a cosmological additive constant, Λ which acts as a repulsive force that prevents the collapse of the Universe. After the publication of the theory of an expanding universe by Friedmann, Einstein realized that his Universe was unstable and even a small change would cause its expansion or collapse. To better

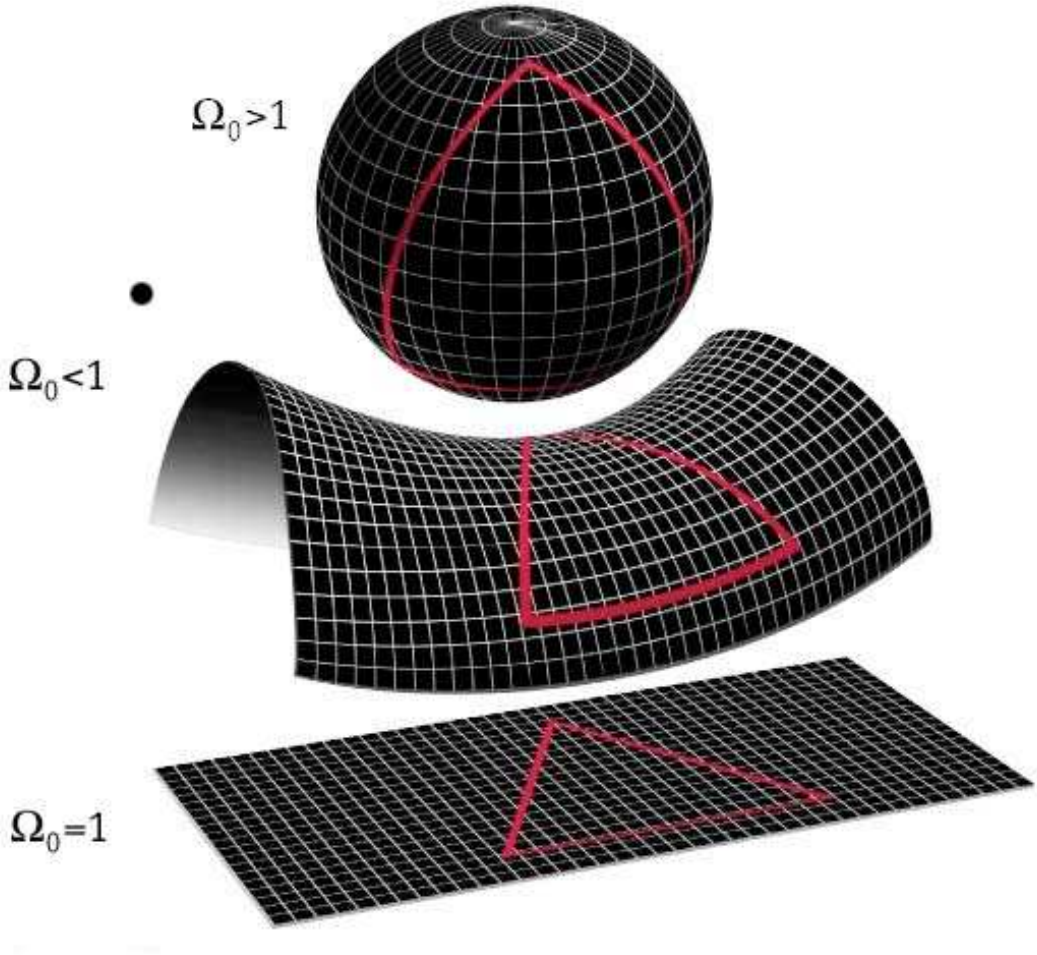


Figure 1.1: Shape of Universe.
From top to bottom: a spherical universe, a hyperbolic universe, and a flat universe.
In a **spherical** universe a positively curved universe is described by spherical geometry, and can be thought of as a three-dimensional hyperspher.
A **hyperbolic** universe is described by hyperbolic geometry, and can be thought of locally as a three-dimensional analog of an infinitely extended saddle shape. For hyperbolic local geometry, many of the possible three-dimensional spaces are informally called horn topologies, so called because of the shape of the pseudosphere, a canonical model of hyperbolic geometry. In a **flat** universe, all of the local curvature and local geometry is flat. It is generally assumed that it is described by an Euclidean geometry, however there are some spatial geometries which are flat and bounded in one or more directions. A positively curved universe is described by spherical geometry, and can be thought of as a three-dimensional hyperspher. You can find this image on http://map.gsfc.nasa.gov/universe/uni_shape.html.

understand the Λ's effects is possible to refer to dynamics equation. Just by taking in to account the simplest case [14], namely when the pressure is zero, it possible to write

$$\frac{1}{a}\frac{d^2a}{dt^2} = -\frac{4\pi G}{3}\rho + \frac{\Lambda}{3} \tag{1.14}$$

in the equation 1.14, a positive value Λ implies a positive contribution for $\ddot{a}$; therefore, Λ acts as a repulsive force. In particular, for large values, Λ becomes larger than the first term, which represents the gravitational force and leads to an Universe acceleration.
In presence of Λ, Ω becomes:

$$\Omega_\Lambda = \frac{\Lambda}{3H^2(t)} \tag{1.15}$$

Note that it is also possible to introduce a dimensionless measure of Λ as

$$\lambda = \frac{\Lambda}{3{H_0}^2}. \tag{1.16}$$

1.1.6 Deceleration

Beside the cosmic acceleration parameter H_0, it is possible to define a *deceleration parameter* that describes the capability of the Universe to react at the expansion.

$$q_0 = -H_0^{-2}\left(\frac{d^2a}{dt^2}\right)_{t_0} \text{ where } H_0 = \frac{\dot{a_0}}{a_0} \tag{1.17}$$

Observing that, H_0 is represents the slope of the relationship between velocity and distance, expressed by the Hubble law (see 1.3), it is possible to say that while q_0 represents the deviation by from the linearity or, in other words, its curvature due to the gravitational deceleration of the cosmic expansion. Moreover, If the cosmic pressure is zero, it is possible to rewrite the density parameter Ω, as function of the deceleration one, as:

$$q_0 = \frac{1}{2}\Omega + \lambda \tag{1.18}$$

where λ is given from 1.16. A universe with $\Omega_0 = 1$ and $\Lambda = 0$ (*Einstein-de Sitter Universe*) has $q_0 = \frac{1}{2}$ [11].

1.2 Standard Model

Cosmological models matching experimental observations confirm that the Universe originated from a singular configuration characterized by infinite density and temperature that still expand itself following the Hubble's law. In the common language this paradigm (will all subtle varieties)is known as: "Big Bang"[3].
With the terms "standard model" or "Big Bang model" it is meant an homogenous an isotropic universe (which fulfills the cosmological principle), with a cosmological constant $\Lambda = 0$, constituted by matter and radiation. In this model, small fluctuations of the density are responsible of the origin for astronomical bodies. Luminous matter is for the most part constituted by Hydrogen and Helium with a density given by:

$$\rho_{0m} = \rho_{0c}\Omega_0 \simeq 1.9 \cdot 10^{-29}\Omega_0 h \ g \ cm^{-3}. \tag{1.19}$$

At the beginning of the Universe, high temperature and density made chemical reactions quite impossible because of the comparable magnitude of reaction velocities and expansion velocity (hereby this kind of Universe is referred as *Hot Universe* (HU). Moreover, in the Universe evolution it is possible to distinguish four different phases in the Universe evolutions:

1. HADRONIC STATE: $t \leq 10^{-6}$ *sec*, $T > 10^{13}K$.
   In this phase the plasma is constituted by neutrinos, mesons, electronic pairs and nucleons-antinucleons couples in thermal equilibrium.

2. LEPTONIC STATE: $10^{-6} < t < 10$ *sec*, $10^{13} < T < 5 \cdot 10^{9}K$.
   In this phase, the only particles able to remain in equilibrium are photons, neutrinos, electron-positron couples, and a few nucleons. This phase is a pure statistical one because the behavior of the particles is determined only by temperature in condition of thermal equilibrium. The matter and radiation density are so high that it is possible for the neutrinos to interact with them. While the expansion proceeds these interaction processes slow down, until such interaction is no longer possible and neutrinos become free to propagate.

3. PHOTONIC PLASMA EPOC: $10 < t < 10^{12} sec, 5{\cdot}10^{9} < T < 4{\cdot}10^{3}K$.

[3]The name Big-Bang was used by supporters of Stationary Models, there after it is become of common use.

It is in this age that the radiation becomes predominant and the annihilation process between electrons and positrons particles can begin. The plasma is ionized due to the high temperature, but it is still in equilibrium with the radiation while the neutrinos are free to propagate. This phase terminate with the hydrogen recombination ($T \simeq 40K$ and $z = \frac{a_0}{a} - 1 \simeq \frac{T}{T_0} \simeq \frac{400}{3} = 1500$).

4. THE AGE FOLLOWING THE HYDROGEN RECOMBINATION: $t \geq 10^{12} sec$.
   When the temperature drops below the range of 4-3 thousand Kelvin, electrons and protons are able to combine to form hydrogen atoms. In this phase, the primary plasma becomes neutral and it is now transparent to radiation.
   The recombination phenomenon has a very high importance since both the perturbations and matter free mean paths increase. In fact before this age, the radiation scattering by photons inhibited the phenomenon. Before the decoupling processes, the matter was offering to the photons larger scattering cross sections (Compton scattering cross sections), however, once hydrogen atoms were formed scattering phenomena almost vanish. In fact the characteristic collision time of a photon with one hydrogen atom is greater than the Universe's expansion time (expressed by $\tau_{0,coll} = \frac{m_p}{\rho_{0m}\sigma_{HC}}$, with σ Thomson scattering cross sections). Moreover, there are still matter density fluctuations until electrons and protons recombine forming electrically neutral atoms able to move freely into the radiation field, thus setting up the seeds for gravitational collapse.

In conclusion, it is possible to say that what in the beginning was a simple fluctuation in an fairly homogeneous universe, ends up into a huge inhomogeneity.

1.3 Structure Formation in different Scenarios

Structure formation scenarios are usually based on the simple principle of gravitational instability. In a medium of uniform density matter, a local density excess (overdensity) will attract nearby matter by the effect of its own gravitational potential.

This type of model predicts a very rapid growth of irregularities. Nevertheless, in the hypothesis of an homogeneous universe, the increase is not so fast as expected. For this reason, it's reasonable to think that there are two effects in competition: on the one hand, there is an increase of the perturbations characterized by gravitational collapse; on the other hand, there is the expansion which tends to dilute all local overdensities. The result is that the growth of fluctuations in an expanding Universe is slowed down.

1.3.1 The Jeans Instability

The first mathematical model for the instability of a homogeneous distribution of matter was formulated by Jeans. This theory takes into account two factors: gravity, which forces the matter to condense, and pressure, which tends to reduce the instability.
Jeans observed that the competition between the two terms is similar to the competition between pressure and gravitation, characterizing the behavior of sound waves (indicating with c_s the velocity of sound) in a collisional fluid where the pressure causes the propagation of sound waves of constant amplitude, whit the gravitation increasing this amplitude.
Furthermore, Jeans noted that a gas cloud can became unstable under certain conditions and it can collapse when the pressure cannot balance the gravitational effects. When the critic mass is reached, the collapse process starts and goes ahead until the forces capable to contrast the process becomes dominant. Jeans formulated a mathematical model to calculate the maximum value of critical mass, as a function of its density and temperature. Moreover, the size of system that sets a limit for the stability was calculated as well.
Taking into account the Hydrodynamic equation:

$$\begin{aligned} \frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u}\cdot\bigtriangledown)\mathbf{u} &= \bigtriangledown\Phi - \frac{1}{\rho}\bigtriangledown p \\ \frac{\partial \rho}{\partial t} + \bigtriangledown\cdot\mathbf{u} &= 0 \\ \bigtriangledown^2\Phi &= 4\pi G\rho \end{aligned} \tag{1.20}$$

where $p, \rho, \mathbf{u}$ and Φ are respectively the unperturbed pressure, density,

velocity and gravitational potential, we can obtain for $\delta = \frac{\delta\rho}{\rho}$

$$\ddot{\delta} = 4\pi G\rho\delta + {c_s}^2 \bigtriangledown^2 \delta.^4 \tag{1.21}$$

The qualitative behavior described above is thus easily obtained: in absence of the pressure term, δ grows exponentially, producing the collapse of clouds in a time:

$$t_c \approx \frac{1}{\sqrt{G\rho}}$$

Instead, taking into account the pressure term, the sound waves take a time t_s to cross the cloud:

$$t_s = \frac{L}{c_s}$$

after this time interval, the system returns to its equilibrium state. However, when $t_c < t_s$, the gravitational contribution becomes larger than the pressure term and the system collapses. The limit of which gravitational instability appears is thus given by $t_s \approx t_c$, which gives

$$L_J \approx c_s G\rho^{-\frac{1}{2}} \tag{1.22}$$

where L_J is the so called *Jeans length* and the mass contaned in a radius L_J

$$M_J \approx {L_J}^3\rho \tag{1.23}$$

is called the *Jeans mass*. This is the minimum mass of a system for which the pressure cannot counter balance the increase in the density contrast δ. As a consequence, a matter distribution having several inhomogeneities will condense into a discrete objects with a size greater than or equal M_J. It is important to observe that for masses comparable with M_J, the pressure in not negligible and the system formed will have, a more or less a spherical shape.

1.3.2 The origin of the matter fluctuations

As already mentioned, during the initial instants of the Universe both density and temperature were extremely high; therefore, the matter was both highly ionized and quickly expanding. In this scenario, all the present

[4]The unperturbed solution corresponding to a static and uniform fluid with $\rho = constant$, $p = constant$

particles, including the photons, are in equilibrium with themselves, hence it is possible to say that the very early Universe behaved as an opaque body. Thereafter, due to the expansion, the system started to cool down thus allowing the matter to decouple from the photons (the temperature dropped below $\sim 10^{-4}$), leading to the beginning of the *recombination epoch*.
After, the photons acquired freedom of movement and the universe became more transparent. These primordial photons are still detected as Cosmic Background Radiation (hereby CBR), characterized by an attenuation factor of ~ 1000.
The CBR spectrum of perturbations was measured in 1989 when the NASA launched in orbit the satellite COBE (COsmic Background Explorer). The results of this mission confirmed that the background radiation has a spectrum of a black body at the temperature of 2.7 K with very small spatial variation in the emission spectrum (order of magnitude of 0.001%) Since the regions of the universe characterized by higher density cooled down more slowly, it is possible to affirm that the CBR temperature gradient gives like a *snapshot* of the matter distribution in the universe at an age of $\sim 10^7$ years. For a temperature of $\delta T/T \leq 10^{-5}$ the cosmic radiation is homogeneous; therefore at the recombination epoch the baryonic matter had the same grade of inhomogeneity

$$\frac{\delta T}{T} = \frac{2}{3}\frac{\delta \rho}{\rho}.$$

Moreover, it is possible to affirm that the CBR provides a direct measurement of the primordial perturbations which by increasing out to gravitational instability, gave birth to the present structures.
As it was already mentioned, the major problem is that on the small scale the Universe presents itself as inhomogeneous while, for increasing scale, this property progressively diminish until the Universe becomes homogeneous on scales larger than 100 Mpc. It is not possible to understand this relationship without first giving a classification for all possible kinds of perturbation; it is also needed to clarify the effects that each kind of perturbation imprinted on the structures. Moreover, before going ahead with this classification it is necessary to underline that is not possible to be sure about the exact shape of the perturbations spectrum, even though it seems logic to assume that it follows a power law.

We can distinguish between two kinds of perturbations,

- **Isothermal** or **entropic fluctuations**: they are associated with a variation of the baryonic density $\delta n_B \neq 0$ in a uniform fluid of photons namely $\delta n_\gamma = 0$, $T = constant$, therefore, the entropy variation s is given by
$$\frac{\delta s}{s} = \frac{\delta n_B}{n_B}$$

- **Adiabatically fluctuations**: in such fluctuations is the entropy that remains constant; therefore, in this case there are variations in both baryons and photons (and in all the particles that contribute to the entropy variation as well). In other words
$$s = constant \qquad \frac{\delta s}{s} = 3\frac{\delta T}{T} - \frac{\delta n_B}{n_B} = 0 \tag{1.24}$$
means that to each density fluctuation it is possible to associate a fluctuation for the CBR temperature
$$\frac{\delta T}{T} = \frac{1}{3}\frac{\delta \rho}{\rho} = 0$$

According to more recent studies, it is possible to affirm that the ratio n_γ/n_B is a constant with a value of $\approx 10^{11}$, hence, given this hypothesis, it is possible to state that the fluctuations that give birth to the cosmic structures were of the second kind (adiabatic), because isotherms fluctuations cannot preserve a constant value for the ratio n_γ/n_B. Moreover the repartition of the fluctuation at a given time, results random and homogeneous in space, hence the *distribution of density contrast* is:

$$\delta(\mathbf{x}, t) = \delta \equiv \frac{(\rho - \bar{\rho})}{\bar{\rho}} \tag{1.25}$$

From which it is evident that it is a function of the random variable $\mathbf{x}$ with a mean value of $< \delta > = 0$ and a correlation function expressed by:

$$< \delta(\mathbf{x})\delta(x + \mathbf{r}) > = \int \delta(\mathbf{x})\delta(x + \mathbf{r}) d^3x \tag{1.26}$$

where $< \delta^2(\mathbf{x}) >$ represents the measurement of the fluctuation in the point $\mathbf{r}$.

The Fourier transform of δ gives as result the wave number k related to the scale L:

$$\delta_k = \int \delta(\mathbf{x}, t) e^{ik\cdot x} d^3\mathbf{x} \tag{1.27}$$

where $|k| = 2\pi R(t)$ is the wave number. It is worth to highlight, once again, that the Eq. 1.27 clearly depends on the scale. In fact performing a choice for the scale as $L \approx 1/k$, and assumed known and fixed the mean value of the density, the initial measurement of the fluctuation is

$$\delta M = \left(\left\langle \frac{\delta M}{M} \right\rangle^2 \right)^{1/2} = \delta = \left\langle \left(\frac{\delta\rho}{\rho} \right)^2 \right\rangle^{1/2} \tag{1.28}$$

The above equation is in relationship with $|\delta_k| = k^n$ form the equation:

$$\delta M \propto \left(\int |\delta_k|^2 k^2 dk \right) \propto k^3 |\delta|^2 \tag{1.29}$$

and furthermore:

$$\delta = \delta M \propto k^{(n+3)/2} \propto L^{-(n+3/2)} \propto M^{-\alpha} \tag{1.30}$$

where $\alpha = (1/2) + (n/6)$. The superimposed value of α is consistent with the existing models since, being smaller than 1, it explains why on the small scale the universe has small interferences, and being larger than zero it ensures the existence of interferences on the large scale.

1.4 The Evolution of Perturbations

It is possible to distinguish between two large classes of possible candidates for Dark Matter: Hot Dark Matter (HDM) and Cold Dark Matter (CDM). The existence of different kinds depends upon a different behavior at the recombination epoch.
As already mentioned, each component of the primordial plasma was in equilibrium with the others, until the characteristic time (peculiarity of each chemical reaction) remains smaller than the Universe expansion time. If this is not the case, decoupling takes place and the number of particles per unit comoving volume is like “frozen”. If this occurs while the particles are still relativistic, the species is said to be *hot*; when they are nonrelativistic, the species is defined as *cold*. In the next sections, are shortly outlined the two

possible scenarios for the evolution of fluctuations

1.4.1 Scenario *Hot Dark Matter*

In this case, the interaction between the baryons and photons tend to dump the oscillatory effect of the fluctuations, under a minim scale, fixed by the Jeans parameter. On the one hand the massive particles tend to assemble together, because of the gravitational potential; on the other hand, the photons, due to their radiation pressure, are able to fight the contraction. In fact, in order for a perturbations of scale L to survive, it is necessary that photons cross it in a time t greater than the expansion time. The adiabatic fluctuations spectrum, at the end of the recombination epoch, is then characterized by a cutoff on the small scale. In the time after the recombination, the perturbations will achieve a mass ($M \approx 10^{13} M_{\odot}$) many orders of magnitude greater than the Jeans Mass ($M_J \approx 10^6 M_{\odot}$). In the HDM models, the formation of structures leads to a top-down formation scenario, in which supercluster of galaxies are the first objects to form after the big bang, while clusters and galaxies form through a subsequent process of fragmentation. This is called also Pancake Model.

1.4.2 Scenario *Cold Dark Matter*

In this case, particles are not relativistic and there is no fluctuation damping, like in the previous one. In this scenario, the first object formed are subgalactic (globular clusters, dwarf galaxies, and so on), while the largest structures form later through gravitational grouping. This is also called hierarchical, or "bottom-up", scenario (*Hierarchical Scenario*).
In the case of isothermal baryons fluctuations, there is a scale dependency, but there is not a cutoff on the small scale. The value of this scale which the initial fluctuation, δM_i, becomes not linear is (with $\Omega_0 = 1$)

$$\frac{\delta M}{M} = \left(\frac{\delta M}{M}\right)_i t^{2/3} = \left(\frac{M_r ec}{M}\right)^{\alpha} (1+z)^{-1} \approx 1. \tag{1.31}$$

1.5 Groups and Clusters of galaxies

In the Universe, the objects tend not to be isolated: stars form stellar clusters which form galaxies; galaxies themselves form groups, groups clus-

ters and so on.
Even our own galaxy, the Milky Way, is a member of the Local Group. This structure was described for the first time by Hubble, at the time of the first distance determinations and redshift measurements (see, his The Realm of the Nebulae,1936). In addition to the Milky Way system, Hubble assigned to the Local Group two satellites, the Large Magellanic Cloud (LMC) and Small Magellanic Cloud (SMC) of our Galaxy, the Andromeda galaxy, (M31, in the Messier' catalog), the Triangulum galaxy (M33). Taking into account recent discoveries of faint dwarf irregulars (dIr) and spheroidal dwarf galaxies (dSph) the total number of presently know members of the Local Group up to ~ 50.
Membership is not certain for all these galaxies, and there are other possible yet unknown members. The most luminous Group members are spiral galaxies. The Milky Way system contains many dwarf galaxies, spread all over the sky, namely Sag DEG , LMC, SMC, the recently discovered and widely disrupted Canis Major, and the dwarf galaxies in Ursa Minor, Draco, Carina, Sextans (dwarf), Sculptor, Fornax, Leo I, Leo II, and perhaps the Phoenix Dwarf, and Leo A (which may alternatively belong to the NGC 3109 subgroup), as well as probably 10 or 11 newly discovered known dwarf spheroidal galaxies.
The system of the Andromeda appears grouped around its main galaxy M31, and contains the dwarf ellipticals M32 and M110 as well as fainter and more far-out NGCs 147 and 185, the very faint systems And I, And II, And III and possibly And IV, And V, And VI (Pegasus dwarf), And VII (Cassiopea dwarf), And VIII, And IX, and And X, as well as the newly discovered spheroidal dwarfs And XI, And XII, And XIII, and And XIV. The third-largest galaxy, the Triangulum spiral M33, may or may not be an outlying gravitationally bound companion of M31, but has itself probably the dwarf LGS 3 as a satellite. Also, IC 1613 may perhaps be a member of the M31 subgroup, as may be also members UGCA 86 and UGCA 92.
The luminosity function of the Local Group contains fewer faint dwarfs than the luminosity functions of many rich clusters of galaxies. The difference is too large to be entirely accounted for by observational bias against the incompleteness of the very low surface brightness Local Group members.
Most of the galaxies in the Local Group belong to clumps that are centered on M31 and on our Galaxy. In the future, interactions between the member

galaxies and with the cosmic neighborhood will continue to change the Local Group. Some astronomers speculate that in a not too far away future the two large spirals, our Milky Way and the Andromeda Galaxy, will collide and merge in some distant future, to possibly form a giant elliptical.
In addition, there is evidence that the nearest large cluster of galaxies, the Virgo Cluster, will probably stop our cosmological recession away from it and accelerate the Local Group toward itself so that it will finally fall and merge into this huge cluster of galaxies.
The analysis of the dynamics of the Local Supercluster has provided a picture of how the enormous mass of the Virgo Cluster acts gravitationally on the galaxies and on galaxy groups around it. As a result, the value of the Virgo Cluster mass and the motion of the Local Group with respect to it (the Virgo-centric flow at the location of our small family of galaxies) seem to indicate that the Virgo Cluster is massive enough to slow down and eventually stop our recession from it, and then accelerate the Local Group member galaxies toward the central region of the Local, or Virgo Supercluster of Galaxies. Therefore, the members of the Local Group and of the Virgo Cluster are the laboratories in which the effects of environment on individual objects can be studied in detail. In general, groups and clusters of galaxies are extremely important to understand the cosmic structure formation process. Their properties, like shape, size, velocity dispersion, are important factors in determining galaxy member evolution processes. In the next sections are summarized the properties of group and cluster.

1.5.1 Galaxies groups

With the word "group" we label a number equal to three or more (up to thirty) of distinct galaxies that are bound gravitationally for periods of time $> 10^9$ years or alternatively, that are the product of a common formation event which has taken during the past 10^9 years even though they may not be gravitationally bound at present. Usually, these galaxies are distributed within a radius R, given by

$$R = (0.1 - 1)h^{-1}Mpc. \tag{1.32}$$

Groups are commonly identified by their conventional name or by the coordinates of their brightest element.

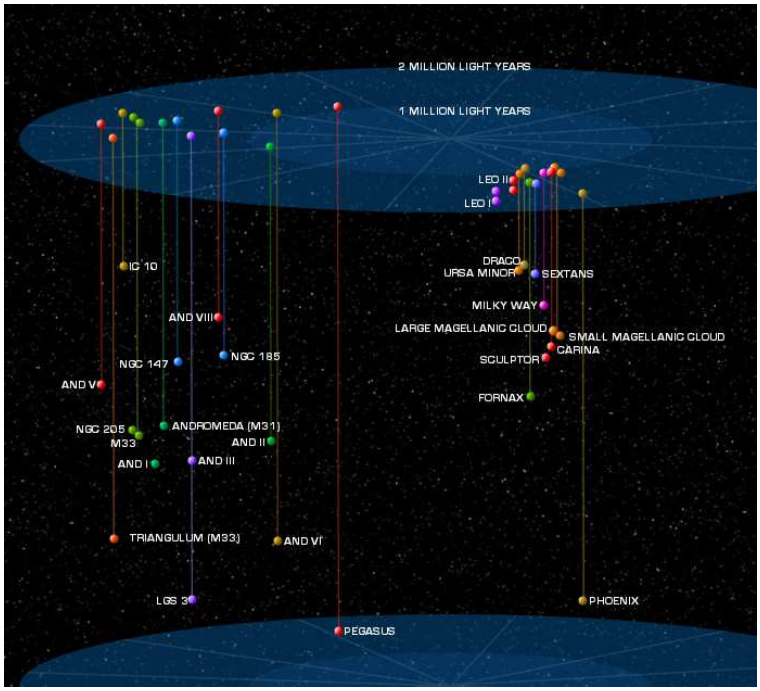


Figure 1.2: **Local Group.**
Schematic illustration of the galaxies locations in the Local Group.

They are usually identified as regions where galaxy density is higher than in the neighborhood regions. However, since it does not exists any supportive physical parameter able to characterize such localization criterion, it is commonly accepted the following definition: an aggregate of galaxies located in a region of the sky, and characterized by a numerical density $\rho_g = N/V$ (N is the number of galaxies present in the volume V taken into account) at least one order of magnitude bigger than the average numerical density of the surrounding volume, can be called a "group".
This definition remarks the fundamental role played by the contrast factor in group identification. An experimental way to confirm that a galaxy is belonging to a group is based on the measurement of relative velocity. In fact within groups such relative velocities can differ only of few hundred $km\ s^{-1}$. However, sometimes, the information on the redshift is not directly available, so in this case to estimate the distance between galaxies it is possible to take in account common characteristics such as, apparent magnitude and diameter.
Nowadays in literature are available several groups classification systems which often are a source of confusion and lead to ambiguous situations. Therefore, it is possible to say that the genetically and or development processes determining the structure of a galaxies cluster and groups are not yet properly understood. However, among the groups it is possible to distinguish between two principal typologies:

1. Compact groups: they are composed by a small number of objects (up to eight) with very high spatial density value;

2. Loose groups: (example: the Local Group) characterized instead by a low spatial density value and large average distance within members ($> 100\ kpc$).

Nowadays it is estimated that about 55% of all galaxies belong to more or less well defined groups.
Additionally, it is possible to mention three further reasons that make the study of groups very interesting. The first one is that such study can bring important clues on the development history of the universe. The second reason is that, the distribution of groups gives information on the matter distribution in the Universe. Finally, studying the physical properties of the interactions within group members and their dependence on the redshift can

also give clues on dark matter and at the same time acts as, on experimental test for cosmological models.
Further details on the physical properties of the groups (in particular regarding the compact ones), are provided in the following paragraphs.

1.5.2 Compact Groups: Morphology

In Compact groups, the separation among members is comparable with their diameters. Therefore the spatial density value is very high, sometimes higher than in the core regions of richest clusters. For example, the 100 compact groups described in the Hicksons catalogue [9] represents about the 1% of the matter in the nearby Universe.
Loose groups are different from compact ones. Firstly, for compact groups the numerical density can be $\sim$ 100 times that of the surrounding space while for loose groups the numerical density can be up to $\sim$ 10 times higher. It is important to observe that the morphology of galaxies within groups varies with their density. In particular, the higher is the density, the lower is the percentage of spiral galaxies. More in detail the decrease of the spiral galaxies fraction in compact groups is due to a decrease in the fraction of late-type ones, while early-type one are still present in groups with high density ($\rho \approx 100h^3\ Mpc^3$). In compact groups the fraction of elliptical galaxies is so high that the spiral fraction (f_s) drops from about $f_s \simeq 0.80$ (for field galaxies) to about $f_s \simeq 0.49$ (Kindl ed Huchra 1988). Opposite is the distribution for loose groups where the distribution is similar to that of the field. Another difference observed between loose and compact group is in their velocity dispersions. Loose groups present a value of about $\sigma_v \approx 200 kms^{-1}$ while compact groups present a value of about $\sigma_v \approx 1000 kms^{-1}$.
Moreover, about the 60% of the spiral galaxies inside compact groups present peculiar rotation curves. It is possible to think that this is a symptom of merging, in fact for few of them it is been observed in a counter-rotation between the stellar and gaseous components of the galaxy. This can be due to the gravitational interaction between two galaxies with different inertial mass and misaligned angular momentum.
Compact groups were pretty unknown till the early 80's. It was only in 1982 that Hickson [9] started a survey on all the available POSS (*Palomar Observatory Sky Survey*) films realizing the first statistically significant catalogue of compact groups using the following three criteria:

1. Population: at least four members ($N \geq 4$)
2. Loneliness: $\Theta_N \geq 3\Theta_G$;
3. Compactness: $\bar{\mu}_G \leq 26.0$;

where N represents the total number of members having a total magnitude in a range of three magnitudes from the brightest member; $\bar{\mu}_G$ is the total magnitude of such galaxies per $arcsec^2$ averaged on the smallest circle (angular diameter Θ_G) containing their geometrical centres; Θ_N is instead the angular diameter of the largest circle that while concentric with the previous one, does not contains any galaxy external to the group with a magnitude neither in the range already defined, nor brighter. The 100 compact groups he found are known as *Hikson compact groups*. They got an average redshift of 0.02 (in other words, recession velocity $\sim 6000\ Km\ s^{-1}$) and so they have a distance between 60 and 100 Mpc.
Before concluding this short introduction it is important to highlight a discovery about the ratio between virial and luminous matter in the groups. In fact the relative fraction of luminous matter of many groups is much smaller than the virial one (Zwicky (1933), Smith (1936), Burbidge & Sargent (1971)). Such discovery made the scientific community think that such groups were not bound systems. However, further observations of the same matter discrepancy in clusters and the discovery of galactic mass halos seems to contradict such hypothesis. It is now widely accepted that groups are systems kept together by their internally dark matter halos. According to Rubin et al. (1991) about the 85% of the mass contained inside compact groups should be dark matter.

1.5.3 Dynamical state of compact groups

Compact groups are characterized by very short dynamic time scales, with $t_{cr} \leq 0.02$, *crossing time*[5], therefore:

1. Compact groups are subject to collapse in a time much shorter than τ_H, this means that the nowadays observable ones are young systems. However, the large number of recognized compact groups is difficult to explain due to the lack of recent or ongoing mergers;

[5]The *crossing time* t_{cr} represents the time needed for a galaxy to cross the group. A low crossing time it is considerate as clue of a real physical interaction.

2. According to Mamon G.A. (1986, [12]) many CGs could be casual alignment of unrelated galaxies within larger groups.

In regard to the listed bullets, in particular about the first one, it is possible to say that the presence of neighborhood galaxies tends to slow down the merging processes increasing the average life span. However, Vennik et al. (1993), by studying the galaxies distribution surrounding some compact groups find that 1/3 of them were part of largest groups.
Therefore, it is likely that the nowadays observable compact groups are young structures. Such hypothesis is also supported by: the absence of primordial gas inside the groups; the low number of interacting neighborhood galaxies; and the little amount of merging residuals. Thus it seems to be confirmed the (Diaferio, Geller and Ramella, 1994 and 1995) hypothesis that compact groups are in continuous formation inside larger groups during their collapse, and so they are a configuration with a short life span.
A theory that takes into account all the mentioned evidences is known as "second merger scenario" (Governato et al. 1996, [8]). According to this theory, a first generation of galaxies developed through a series of merging (first infall), into early type and very bright galaxies nowadays present inside the groups. Then the galaxies present outside this overdensity region once reached the inversion point, started to fall (secondary infall) towards the central region. As a conclusion, it is the second infall that provides to the groups new members and gives the present aspect.
There is also a difference in the intergalactic space composition between the two generations. In fact, in the second generation this space is filled from matter stripped from the primordial galaxies that, by reducing the scattering sections increases the average life span of the groups. This observation also gives a possible explanation for the high number of bright elliptical galaxies (Caon et al. 1994) that seems to dominate the dynamic of the groups (Shaker et al. 1998).
Results of studies on compact groups using several models of the universe show that their population is different and moreover, it is in relationship with age and type of the universe model (critic or open). However, it is close to zero the dependence from the original fluctuation type.
In a critic universe model, quite all the observable groups went through a strong interaction and merging phase during the collapse of the central region characterized by a redshift within a range of ~ 0.35 and ~ 1. Thereby,

according to Governato et al. (1996) the merging ratio decreases, therefore, a large fraction of the currently visible groups are result of the first infall. Afterword galaxies lose their characteristics associated to merger events giving a theoretical justification of the low fraction of bright blue elliptical galaxies within groups. Thus it is possible to say that in a critic universe the number of expected groups results coherent with the observed one.
In an open universe models instead, the largest part of the clustering process is supposed to be at high redshift values ($z > 0.3$) followed by a strong decrease of the formation rate. Therefore, in the open universe model the secondary infall phase is absent; hence, the observed group number results greater than the expected one. Moreover, in order to consider a low density universe, the galaxies numerical density should results lower than the observed one.

1.5.4 Triplets of galaxies

Groups with only three members have high importance in the group studies because they are the conjunction element between binary galaxies (that can be described as an interaction between two bodies) and largest groups with many members. Despite their importance, since that it is really difficult to have a complete catalogue, there is a lack of literature on the triplexes. However, a catalogue of triplexes is available; as a result of a study conducted at the end of the '70s by the Russian astronomer Igor Karachentsev. Members of this catalogue are known as K-triplets. Later studies discovered that more than half of them are physically bound systems and not projective effects. Some important parameters on the triplets are now known such as:

- The value of the dispersion velocity:

$$V = \left[\frac{1}{3}\sum_{i=1}^{3}(V_i - < V >)^2\right]^{1/2} = 101\ km\ s^{-1}$$

  where V_i represents the radial velocity of the ith galaxy inside the triplets and $< V >$ is the average of the radial velocity.

- The average dimension: $R \simeq 43\ Kpc$;

- The luminosity mass: $M_L \simeq 3.8 \times 10^{11}\ M_\odot$;

- The Virial Mass: $M_{VT} \simeq \alpha \frac{V^2 R}{G} = 1.7 \times 10^{12}\ M_{\odot}$
  where α is a dimensionless parameter linked to the projection effect of V and R, in this case its value is assumed to be $\alpha = \frac{3\pi}{2}$.

- The characteristics *crossing time*: $t_{cr} = 2H\frac{R}{V} \cong 8 \times 10^8\ years$

Given the previous relationship, it is possible to derive the mass-luminosity ratio

$$\frac{M_{VT}}{M_L} \cong 4.6;$$

Moreover, according to the results obtained by Zheng & al. (1993) [16], triplets seem to originate from merging of galaxies belonging to larger groups, and leading to the formation of elliptical ones (about the 44% of the galaxies present in the model seem to be merging resituate). This hypothesis is supported by several facts, because in such groups it is not present a fraction of early type galaxies (E and or SO) larger than in the field. Based on the possible four apparent positions that the three members can assume (these are already known as projections), Agekian & Anosova (1967) [1] gave to each one a proper name: Lagrangian (**L**), Hierarchical (**H**), Aligned (**A**) and Median (**M**).
Experimental observations seem to confirm a quasi uniform distribution for the recognized triplexes among the four possible configurations. However, there is just a bit of excess for the A configuration, compensated by a slightly defect for the M configuration. It is worth to highlight that further studies conducted form Chernin (1994) on the galaxy evolution, report a higher concentration of configuration **H**. Such high concentration brings forth the hypothesis that there is a deficit of such configuration which is not possible to explain with a simple projection effect.
It is worth noticing that an uniform triplet distribution, once projected on the sky, is not anymore uniform, but presents an excess of H configuration and this supports the hypothesis that such configuration has not a long lasting life span. In other words, for such configuration a merging would occur each time that two members come close. Such hypothesis was not taken into account by Chernin. However, such high merger frequency would result into an increase of the number of elliptical galaxies that is not supported (as mentioned) by observations. In conclusion, it is possible to affirm that the merging process has not very much influence on the development and life of triplets.

A possible explanation could come from the presence of dark matter. In fact, the interaction between members of an H configuration (two galaxies are closer) is in prevalence a gravitational attraction that will bring at some stage the two galaxies to be very close to each other.
However, the presence of diffuse dark matter inside the group could weaken the gravitational attraction contributing to increase the number of **H** configuration expected by Chernin in the hierarchical zone.
On the other hand in a bottom-up scenario (or hierarchical), the larger structures originates from the gravitational attraction between smaller ones. In such scenario structures members of a system could have independent origin, and it is not possible to explain a non casual and uniform distribution in the orientation of the galactic planes.

1.5.5 Cluster

A collection of at least thirty galaxies gravitationally interacting and contained in a radius $R_c = 1.5h^{-1}Mpc$ (known as *Abell's Radius*), can be called *cluster* if the magnitude of each member satisfies the relationship:

$$m_g \leq m_3 + 2^m$$

where m_3 is the magnitude of the third brightest galaxy in the cluster (Abell 1958). Clusters are characterized by a spatial density which on average is lower than the groups. However they have relaxation time at least one order of magnitude smaller than the Hubble time. The other more important characteristics of groups and clusters are summarized in Tab.1.1

It is important to clarify that our knowledge on the Universe at the largest scales is limited by our ability to observe and to measure the distance of distant away objects. However, clusters are relatively easy to recognize and it is also relatively easy to estimate the distance between members. Clusters are divided in two categories based on the presence or absence of central symmetry. Clusters are defined *open or irregulars* when it is not possible to recognize a central region (e.g. the Virgo one). Alternatively, they are classified as *spheroidal* ones when they present a more compact spherical symmetry and a central region with higher density than the more external one. The number of galaxies present within the Abell radius of a cluster, it is called *richness* ($\mathcal{R}$). The integral density ($n_c > N_{\mathcal{R}}$, a value of

Property	**Groups**	**Clusters**
$\mathcal{R}$	$3 - 30$ galaxies	$30 - 30$ galaxies
R	$(0.1 - 1)h^{-1}Mpc$	$(1 - 2)h^{-1}Mpc$
σ_V	$100 \div 500 kms^{-1}$	$400 \div 1400 kms^{-1}$
m	$(10^{12.5} - 10^{14})h^{-1}M_\odot$	$(10^{14} - 2 \times 10^{15})h^{-1}M_\odot$
L_B $(r \leq 1.5h^{-1}Mpc)$	$(10^{10.5} - 10^{12})h^{-2}L_\odot$	$(6 \times 10^{11} - 6 \times 10^{12})h^{-2}L_\odot$
$< M/L_B >$	$\sim 200hM_\odot/L_\odot$	$\sim 300hM_\odot/L_\odot$
$\bar{n}$	$(10^{-3} - 10^{-5})h^3Mpc{-}3$	$(10^{-5} - 10^{-6})h^3Mpc{-}3$

Table 1.1: **Typical property of groups and Cluster.**
$\mathcal{R}$ is the *Richness*, it is the number of galaxies into a structure. With L_B mean the typical blue luminosity R is the mean radius. σ_V is the dispersion of radial velocity. m denotes groups mass. $< M/L_B >$ is mass-luminosity ratio[6] and $\bar{n}$ is numerical density.

density over a certain richness threshold) and the mean separation ($d \equiv \bar{n}_{1/3}$, where $\bar{n}$ is the average density across clusters) are summarized across clusters in Table 1.2.

$\mathcal{R}$	$N_{\mathcal{R}}$	$n_c(> N_{\mathcal{R}})h^3 Mpc^{-3}$	$d(> N_{\mathcal{R}})h^1 Mpc^{-3}$
≥ 0	≥ 30	13.5×10^{-6}	42
≥ 1	≥ 50	6.0×10^{-6}	55
≥ 2	≥ 80	1.2×10^{-6}	94
≥ 3	≥ 310	1.5×10^{-7}	188

Table 1.2: **Numerical density of cluster**

Clusters, in particular the richer ones ($R \geq 100$) are not common objects; they are characterized by a density $\leq 10^5$ per clusters Mpc^3. They are also characterized by an high average velocity dispersion $\sigma_V \sim 750\ km\ s^{-1}$ (in range $\sigma_V \sim 400 \div 1400\ km\ s^1$).

Before concluding this short introduction it is worth to underline the weak, but still present relationship between σ_V and the richness: richer clusters have larger σ_V.
Moreover, observation in the X ray-band shown that in all richer clusters, together with the galaxies it is present also hot intergalactic plasma with a temperature within the range of $\sim 2\ kev$ and $\sim 14\ kev$.

1.5.6 Morphology of clusters

In less rich clusters a big fraction of the members is composed by spiral galaxies. In rich ones instead, the fraction of $S0$ and elliptical one increases towards the central region while the spiral fraction decreases. It has been observed that in few of the richest clusters the fraction of spiral is close to zero in the core region.
According to *Spitzer & Baade* (1951), a possible explanation for the lower count of spiral galaxies in the core region of rich clusters could be the result of a number of collisions between galaxies across the core region (from the birth of the Universe) that stripped out their stellar material. However, due to the latest correction in the estimated dimensions of the clusters, this hypothesis would never be satisfied. In fact is been proved that this hypothesis would be satisfied only if the estimated age of the cluster results

larger than the estimated age of the Universe. Hence, it is possible to say that spiral galaxies inside rich clusters, either have never been present or have disappeared by an unknown development process.

1.5.7 Mass, luminosity and mass-luminosity ratio

Since there is a relationship between the critical mass Ω and the mass of the cluster, the determination of the mass is an appealing field of study in cosmology. For a single cluster the approximate range for its mass is:

$$M_{cl}(R \leq 1.5)(0.12)10^{15}h^1 M_\odot. \tag{1.33}$$

The typical (median) blue luminosity of rich clusters (within 1.5 h^-1 Mpc) is:

$$L_{cl}(\leq 1.5) \sim 10^{12}\ h^{-2}\ L_\odot$$

so the approximate range of rich cluster blue luminosities is

$$L_{cl}(\leq 1.5) \sim (0.6 - 6) \cdot 10^{12} h^{-2}\ L_\odot.$$

Therefore the typical mass-to-luminosity ratio of rich clusters is thus

$$(M/L_B)_{cl} \sim 300h(M_\odot/L_\odot).$$

Such values confirm the presence of dark matter in huge quantities. From further studies on the cluster dynamics a density parameter for the Universe has been estimated as:

$$\Omega_{din} \sim 0.2 \qquad \text{if} \qquad M \propto L \qquad \text{on a scale} \qquad \geq 1h^{-1}Mpc.$$

At the value $\Omega = 1$, that is the critical value in order to consider the Universe closed, it is possible to write:

$$(M/L_B)_{(\Omega=1)} \simeq 1500h.$$

Due to the presence of dark matter, the mass-to-luminosity ratio grows out to a radius the order of $(0.1 - 0.2)h^{-1}$ Mpc, then slows down and remains almost constant either in groups, either in clusters. The plateau value is $M/L_B \simeq (200 \div 300)$ h (That correspond to a $\Omega \simeq 0.2$). Hence, the largest

fraction of dark matter present within cluster is due to the contribution from the galaxies halos and to the plasma present inside and around the individual galaxy members. Therefore, there not be should additional amount of dark matter within clusters, beside such mentioned one.
Unless the mass distribution in cluster is totally different from that of the luminous matter with an huge presence of dark matter in the void or on the very large scale, the measured universe density should be much smaller (~ 0.2) than the critical one.

1.5.8 Mass density estimation: Ω

Rich clusters of galaxies provide the best laboratory for studying the baryon fraction (i.e., the ratio of the mass in baryons to the total mass of the system) on relatively large scales.
In the clusters, the baryonic mass is composed by two principal components: the luminous portion of the galaxies and the intergalactic gas. So, the baryon fraction in clusters is

$$\frac{\Omega_b}{\Omega_m} \geq \frac{(M_{gas} - M_{stars})}{M_{cl}} \simeq 0.07h^{-1.5} + 0.05 \tag{1.34}$$

where the first term on the right-hand side represents the gas mass ratio[7] and the second one the fraction of stars. The baryon density required by big-bang nucleosynthesis is

$${\Omega_b}^{BBN} \simeq 0.015h^{-2} \tag{1.35}$$

Please observe that if in the previous relationships $\Omega_m = 1$, the baryons number observed inside the clusters would be greater than the one allowed by the big-bang. Combining equations 1.34 and 1.35 it is possible to estimate

[7]The ratio of the mass of gas in clusters to the total virial cluster mass (within $\sim 1.5h^{-1}Mpc$) is observed to be in the range

$$\frac{M_{gas}}{M_{cl} \simeq (0.03 - 0.15)h^{-1.5}}$$

the total gas mass in clusters is generally larger than the total mass of the luminous parts of the galaxies (especially for low values of h). With so much gas mass, it is most likely that a large fraction of the intracluster gas is of cosmological origin (rather than all the cluster gas being stripped out of galaxies).

the value for the mass density:

$$\Omega_m \leq \frac{\Omega_b}{0.07h^{-1.5}+0.05} \simeq \frac{0.015h^{-2}}{0.07h^{-1.5}+0.05} \sim 0.2$$

with $h \sim 0.5 \div 0.8$. Thus there should be a huge numbers of baryons imprisoned within the dark matter present inside the clusters.

1.5.9 The mass function

The clusters mass function describes the numerical density of clusters characterized by a mass larger than a threshold value M. This is a fundamental quantity to discriminate between different theories on the origin of the Universe. In fact it is believed that larger clusters originated from rare fluctuation peaks in the distribution of the mass density, hence, less rich clusters and groups had birth from smaller amplitude fluctuations. The observed function it is well described by the relationship (Bachall,1996[2]):

$$n(\geq M) = 4 \times 10^{-5}(M/M^*)^{-1}exp(-M/M^*)h^3Mpc^{-3}$$

With $M^* = (1.8 \pm 0.3) \times 10^{14}h^{-1}M_\odot$.
Comparing the observed mass function with the one expected from several cosmological model, it is possible to observe that the standard model Cold Dark Matter (CDM) (with $\Omega = 1$) leads to huge clusters, while to obtain the observed ones, Ω should be $\sim 0.2 \div 0.3$.

Chapter 2

The Multiplicity function

Studying galaxy clustering brings forth important clues on the conditions at the age of recombination. Three are the most commonly used ways to characterize the irregular distribution of both matter and light: the luminosity function (hereby LF), the correlation functions (hereby CF) of various orders and the multiplicity function (hereby MF). These quantities are useful tools in Cosmology because they all describe (in different manners) the cosmic abundance of objects.

2.1 Correlation Function

The probability to find galaxies in a small volume sample (taken from a large homogeneous sample of galaxies Poisson distributed), at a distance r from each other, is defined using the formula $\delta P = n\delta V$, where n is the numerical mean density of galaxies present in the sample. It is worth to clarify that if in the sample are also present structures, the probability must be calculated by the formula:

$$\delta P = n[1 + \xi(r)]\delta V \tag{2.1}$$

In a few words, it is possible to affirm that the the CF allows to calculate the number of couples present around a small volume δV in the sample and furthermore, to compare it with the number of couples presents in a Poisson distribution of galaxies included in the same sample. Hence, assuming N_G as the number of galaxies in the sample, and N_P as the number of casual points in the volume element taken into account, it is possible to obtain $\xi(r)$

from:

$$1 + \delta(r) = \frac{n_{GG}(r)}{n_{RG}(r)} \frac{N_R}{N_G} \tag{2.2}$$

where $n_{GG}(r)$ represents the number of couples with a separation factor r that falls in the range between $r - \Delta r$ and $r + \Delta r$, also $n_{RG}(r)$ represents the number of couples where one of the members is a casually distributed point, while the other one is a galaxy, again with a separation factor that falls in the already specified range. It is therefore possible to infer that the CF provides indications on the tendency of the objects considered to group themselves together in structures. A zero value for the CF ($\xi(r) = 0$) corresponds to an uniform distribution, while for $\xi(r) > 0$ ($\xi(r) < 0$) the distribution is more (or less) skewed. Definite as small scale any values for r that falls in the interval $0.1h^{-1}Mpc \leq r \leq 10h^{-1}Mpc$, the CF follows a power law expressed by the equation:

$$\xi(r) = \left(\frac{r}{r_0}\right)^{\gamma} \tag{2.3}$$

where $\gamma \sim 1.8$ and the correlation length is $r_0 \sim 5h^{-1}Mpc$. Moreover, it is also possible to define $\xi(r)$ thought the *power spectrum*[1]

$$\xi(r) = \frac{1}{2\pi^2} \int dk k^2 P(k) \frac{\sin(kr)}{kr} \tag{2.4}$$

such definition, shows that $\xi(r)$ is linked to the spectral fluctuation of the density at the initial recombination.

2.2 Luminosity Function

The luminosity function (LF) gives the number density of stars or galaxies having luminosity in the range between $(M, M + \Delta M)$. Based on the observation made by Press & Schechter (1974) and later by Schecter (1976), it is possible to state that the LF follows a power law. In particular, the number of galaxies decreases, with a monotonic behavior, while the luminosity increases. For faint magnitude, $\Phi(M)$ decreases with an exponential

[1]The power spectrum P(k) reflects the fluctuation in the amplitude of the density field that gave birth to the existing galaxies and structures. As logic consequence it follows a power law$P(k) \propto k^n$. Usually $n = 1$ in this case, in fact, the fluctuations on different scales are correspondent to the same amplitude fluctuations in the gravitational potential

behavior with respect to $|M|$ and reaches zero for a particular value of the magnitude defined as M^*. Changing the magnitude with the luminosity, moreover, given the definition of $\Phi(L)dL$ as the number of galaxies with the luminosity that falls in the range of $(L, L + \Delta L)$, it is possible to write the LF as:

$$\Phi(L) = \left(\frac{\Phi^*}{\Phi}\right)\left(\frac{L^*}{L}\right)^{\alpha} exp - \left(\frac{L}{L^*}\right) \tag{2.5}$$

where L^* is the brightness that correspond to M^*. In the case groups are close enough to the observer, it is possible to obtain the photometric information for weak objects as well, thanks to the assumption that the distance can be assumed to be the same for all the objects in the same group. However, for the richest and far groups, it is more complicated to obtain such photometric information. According to Bachall (1979), it is possible to attempt to solve the mentioned difficulty by assuming that the total luminosity of a rich cluster increases with the number of its members. Hence, if L_{clu} expresses the luminosity of the cluster and with $n_{gal}(R)$ the number of members as function of the richness factor R, it is possible to write:

$$L_{am} \sim n_g al(R) < L/n >_0 \tag{2.6}$$

where $< L/n >_0$ represents the average brightness. So the LF for groups of galaxies can be written as it follow:

$$\eta(L)dL = \eta_0\left(\frac{L}{L_0}\right)^{-\alpha} exp\left(-\frac{L}{L_0}\right)dL. \tag{2.7}$$

In 2.7 η_0, L_0 and α are unknown parameters. In particular η_0 is an a-dimensional normalization factor; L_0 is the brightness value where the curve $\log \eta$-$logL$ exhibits a marked slope change; α is the LF slope at faint light intensities ($L < L_0$). According to results by Turner & Gott (1976,[15]), the application of 2.7 to small galaxies groups gives $L^* = 3.4 \times 10^{10} L_{\odot}$, corresponding to an absolute magnitude of $M^* = -20.85$.

2.3 Multiplicity Function

The multiplicity function (MF) is used to calculate the amount of objects as function of the mass, or as a function of brightness, or as function of any other arbitrary parameter used to define the objects into account.

2.3.1 MF as function of the brightness

Taking into account a region of the universe large enough to be a statistically significant sample then, let us to compute its average brightness [2]. Selecting inside the region the largest sphere that exhibits internal brightness satisfying the relationship:

$$\rho_{L,sphere} > X_{\rho L} \tag{2.8}$$

where $X_{\rho L}$ is a well chosen contrast factor. The matter present inside the mentioned sphere is the constituent of the first group that has total brightness L_1. Selecting the second largest sphere in the region which satisfies the relationship 2.8, it is possible to build a second group and so on. The recursive application of the mentioned method allows selecting all the groups contained inside the region taken into account. As a result it is possible to obtain a complete catalogue of groups at specified density values. It is now possible to define the MF as the function of groups luminosity, $\eta_g(L)dL$ or, in others words, the number of groups having luminosity in the range between L and $L + dL$. It is also possible to approximate $\eta_g(L)dL$ with $\eta_g(N)dN$ using the relationship $L \approx 3NL^*$, where N is the number of the group members. The mentioned approximation is based on the fact that in a complete survey, until the magnitude does not reach a certain limit value, the mean of the luminosity is expressed by the formula $L = 3L*/2$ [7]. Moreover at the average distance where the galaxies are, about half of the total group luminosity comes from visible components. However, the mentioned approximation has its limitation: the brightness of the closest groups might be over-estimated.

2.3.2 MF as functions of the mass

In a similar way, it is possible to define the MF as function of the mass. In this definition $\eta_g(M)dM$ represents the numerical density for groups with mass value that falls in the range between M and $M + dM$. According to Press & Schechter (1975), the MF follows an exponential law, and in particular it is possible to write:

[2] Turner & Gott (1976) [15] estimated the value of such density as:$\rho_L = 4.7 \times 10^7 L_\odot Mpc^{-3}$

$$\eta_g(M)dM = (2\pi)^{-1/2}\left(\frac{\bar{\rho}}{M_c}\right)(1+(\tfrac{n}{3}))\left(\frac{M}{M_c}\right)^{-(3/2-n/6)} \times exp\left[-(\tfrac{1}{2})\left(\frac{M}{M_c}\right)^{(1+n/3)}\right]\left(\frac{dM}{M_c}\right)$$

where M_c is called cutoff mass (1). It is important to highlight that if the ratio M/L is known for a different class of objects, it becomes possible to link the MF as function of the luminosity and as function of the mass. Additionally, if the ratio M/L is a constant, the MF as function of luminosity and mass are consistent.

2.4 The Integrated MF

Often it is convenient to introduce the integral of the MF, directly on a chart, namely $f(N)$ represents the portion of galaxies inside groups with components $\leq N$. The integrals of MF across catalogues with different magnitude integration limits, leads to different N_{max} values. It represents the maximum number of components inside a group. This is due to the finite number of galaxies present in the sample. Furthermore, the size of the largest group inside the sample is obviously limited by the finite sample size itself.
In fact: *the deeper is the survey, the flatter are the curves of MF corresponding to large N values.* Also, we can express the MF as the integral of the luminosity

$$f_g(L) = \frac{1}{\rho_L}\int_0^L L\eta_g(L)dL \tag{2.9}$$

where f_g is the portion of the total luminosity in the universe included in groups with brightness $\leq L$. This portion is expressed normalized to values between 0 and 1. Recalling the approximation in 2.3.2 it is possible to write:

$$f_g(L) \approx f_g(3NL^*) \tag{2.10}$$

2.5 The importance of the MF

The CF do not give a complete description of the clustering. In fact many different clustering could have the same first order CF but completely

different three-, four-,five-, etc. points CF. Thus it is necessary to have a complete set of CFs to give a complete description of the clustering. However, it is not trivial to obtain CF of higher order: in fact, for a sample containing N galaxies, the $k-th$ order CF requires $\sim n^k$ operations. Ad example the CF of the third order ζ it is definite as cumulative probability to find the three galaxies centered into the three volumes dV_1, dV_2 and dV_3:

$$dP = n^3 dV_1 dV_2 dV_3 [1 + \xi_{12} + \xi_{23} + \xi_{31} + \zeta] \tag{2.11}$$

Additionally, one of the heaviest limitations is that on a scale larger than $10h^{-1}Mpc$, the CF exhibits a fast decreasing behavior, while the errors are proportional to the uncertainty of the numerical density of the galaxies: $\Delta\xi = \frac{\Delta\bar{n}}{\bar{n}}$.
This means that to measure the fluctuations on the largest scales it is much more convenient to use the power spectrum (hereby PS) that reflect the fluctuation amplitude as a function of the scale used. However, the experimental measurement of the PS requires a number of normalizations due to the different nature of the objects used. As a result of the normalizations, a biasing factors b is present:

$$b \equiv \left(\frac{\Delta\rho}{\rho}\right)_{gal} / \left(\frac{\Delta\rho}{\rho}\right)_{m} \tag{2.12}$$

It is possible to say that the bias factor b (expressed above) represents the over-density of the objects taken into account related to the over-density of the mass present into the sample. The MF is complementary to the CF because it gives the ratio between the amplitude of the CF of superior orders and the 2-points function. Hence, the MF is a well built instrument that allows the study the power spectrum of the density fluctuation. In particular the MF provides clues on the index of the initial fluctuations.
The most important advantage of the MF is that it is not so difficult to calculate because once known the group catalogue, it is required to know only the number of galaxies belonging to each shell around each one of the galaxies. Additionally, it is worth to highlight once again that both MF and LF follow the same mathematical rule (see eq.2.5). In fact both follow an exponential bright end power law. However, one problem remains, the LF is not a natural extrapolation of the gravitational clustering phenomenon. In addition, the slope of the average LF correspond to a n value in contrast

with the one that is possible to find using the MF and CF methodology [7]. Another problem in the precise characterization of the MF is the production of a groups and clusters catalogue with a high statistically significance and covering a large portion of the sky. In other words, the catalogue should include structures on a large range of richness (from structures with very low multiplicity, such as triplets of galaxies, up to the largest clusters with hundreds of members).

Chapter 3

Friends Of Friends Algorithm

In this chapter, two different versions of Friends of Friends (hereby, FOF) algorithm are presented. The first is based on the algorithm presented by Huchra & Geller [10], to be used on spectroscopic data. The second one is based on the changes proposed by Botzler et al. [4] for applying the same method to photometric data. Before describing these two different versions, it is important to emphasize the differences between photometric and spectroscopic redshifts. As mentioned in Chapter 1, the redshift of a galaxy is defined by

$$z = \frac{\delta\lambda}{\lambda} = \frac{(\lambda - \lambda_0)}{\lambda_0} \tag{3.1}$$

where λ_0 is the wavelength observed into laboratory frame of reference. So the redshift is the displacement toward lower frequencies of electromagnetic spectrum due to Universe expansion. The measure of $\Delta\lambda$ is an operation which can be complex and expensive in terms of observing time: for each galaxy, in fact, medium-high dispersion spectra must be obtained (few tenths of Angostrom per mm) and so it is necessary to use medium large telescopes and long exposure times. Redshifts can be estimated spectroscopically recognizing absorption or emission lines and their wavelengths (λ) in the spectrum, and measuring their position with respect to the positions that the same lines have in the laboratory frame reference (wavelengths at rest λ_0). This is the reason why until a few years ago, *spectroscopic redshift* were available only for a few thousand galaxies.

Photometric Redshift use instead large band photometry: they are obtained using photometric data in different bands: magnitudes or color indexes. To explain this approach, an example may be needed: macroscopic features of

spectrum (such as the break observed around 4000 Å) of the same galaxy at different distances fall at different wavelengths, and so they are weighted in different ways within each photometric band. The result is that galaxies at different distance have different colors, by inverting the problem it is possible to obtain an estimate of the redshift (the so called photometric redshifts). A measure of photometric redshift is substantially a measure of color and so therefore: it is necessary to distinguish intrinsic colour changes of galaxies from those produced by cosmological effects.
Photometric redshifts are faster and easier to obtain than spectroscopic ones; their measurements provide a method to evaluate distances when it becomes impossible to make spectroscopic estimations due to either poor signal to noise ratios, or the instrumental systematic error. Photometric redshifts have lower accuracy ($\sim \pm 0.02$) than spectroscopic ones ($\sim \pm 0.001$), but if available in large numbers and for statistically well controlled samples, they can provide information on the 3-D distribution of objects of Universe, thus reducing the effects of projection of galaxies due to foreground and background galaxies.

3.1 The Friends Of Friends algorithm

The Friends Of Friends algorithm (hereby, FOF) was created to search for groups in spectroscopic galaxy surveys and has been modified to find structures in simulated galaxy data sets. This algorithm presented by Huchra & Geller has several attractive features: it can be applied to a variety of samples, it is easy to apply and to implement on a computer and it can handle many different selection effects.
The procedure is outlined in Figure 3.1. Firstly, the algorithm chooses a galaxy in the catalog which has not been previously assigned to a group. Then, searches around it for companions with small projected separation. In other words, if i and j are two galaxies with angular separation θ_{ij} and redshift z_i, z_j, they have a projected separation D_{ij} given by:

$$D_{i,j} = 2\sin(\frac{\theta}{2})\frac{\bar{V}}{H_0} \tag{3.2}$$

where V is

$$V = \frac{(V_i + V_j)}{2} \tag{3.3}$$

and their line of sight separation is

$$V = |(V_i - V_j)| \tag{3.4}$$

Now, the algorithm defines i and j "friends" if

$$D_{i,j} \leq D_L(V_i, V_j, m_i, m_j) \tag{3.5}$$

$$V_{i,j} \leq V_L(V_i, V_j, m_i, m_j) \tag{3.6}$$

where m_i, m_j are their magnitudes, D_L and V_L are respectively projected and the line of sight "linking lengths". If both these conditions are not satisfied, (rather no companion of i are found), the galaxy i is entered in a list of "isolated" galaxies. All companions found, are instead added to a list of group members. The neighbors of each companion are then searched. This process is repeated until no further members can be found.
It is important to observe that this procedure is commutative: if a galaxy i is as companion the galaxy j than galaxy j is a companion galaxy i. Commutativity means that for any particular galaxy catalog and choice of selection parameters, a unique group catalog results.

3.1.1 Linking Lengths

The most important ingredients of this group finding algorithm is in the choice of the linking length parameters. In fact, if the linking lengths are too small, then the group finder will break up single structures into several groups. If the linking lengths are too large, then different structures will be merged together into a single one. There are no a priori and optimal criteria to fix these parameters that will work optimally on any catalog. The right choice of linking lengths depends on the purpose for which groups are being identified.
The method adopted by Huchra & Geller was designed specifically to compensate for variations in the sampling of the galaxy luminosity function as a function of the distance of the group (Malmquist bias). They assumed that the LF is independent on the distance and position and that at larger

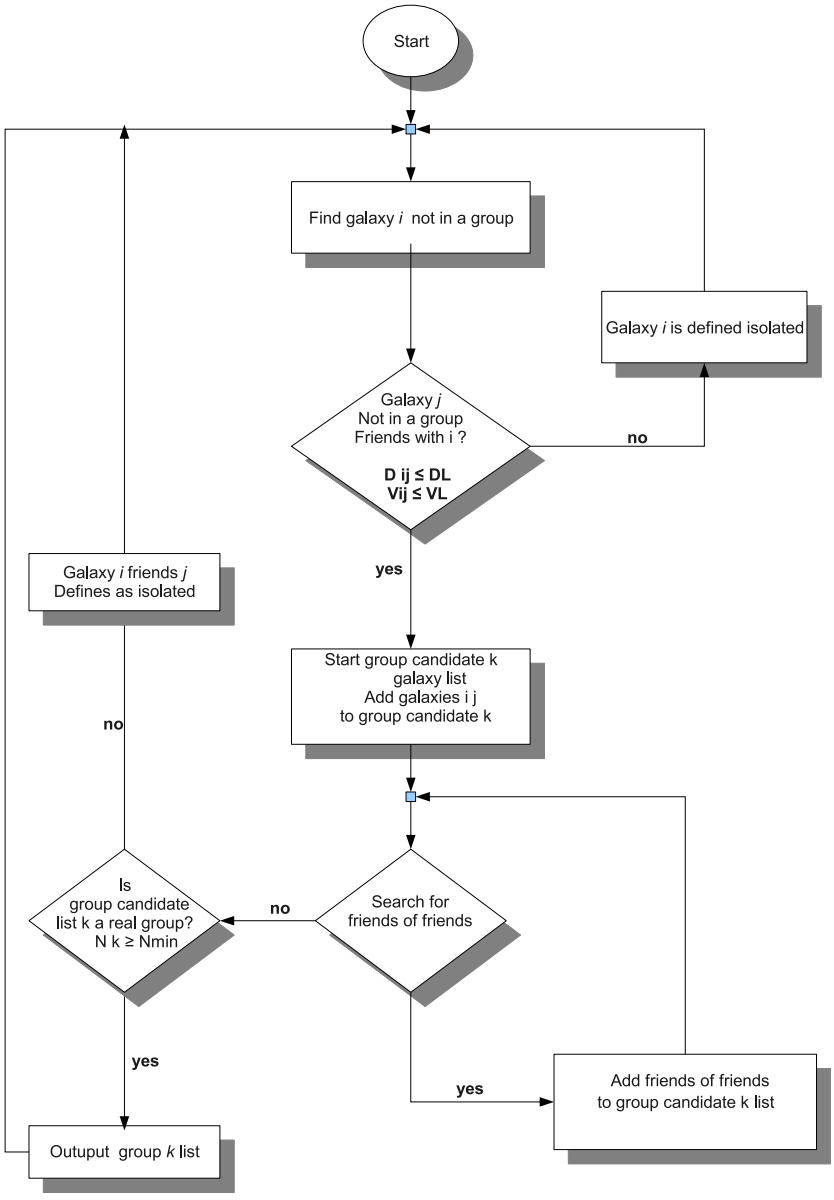


Figure 3.1: Flow chart for group selection algorithm

distances the fainter galaxies are missing. So, they used

$$D_L = D_0 \left[\frac{\int_{-\infty}^{M_{ij}} \Phi(M) dM}{\int_{-\infty}^{M_{lim}} \Phi(M) dM} \right]^{1/3} \tag{3.7}$$

where

$$M_{limit} = m_{limit} - 25 - 5log(V_F/H_0)$$

and

$$M_{ij} = m_{limit} - 25 - 5log(V/H_0)$$

$\Phi(M)$ is the differential galaxy luminosity function[1] for the sample, D_0 is the projected separation in Mpc chosen at some fiducial redshift V_F. Additionally, Huchra and Geller choose the limiting velocity difference as:

$$V_L = V_0 \left[\frac{\int_{-\infty}^{M_{ij}} \Phi(M) dM}{\int_{-\infty}^{M_{lim}} \Phi(M) dM} \right]^{1/3} \tag{3.8}$$

The ratio V_0/D_0 is related to an assumed cosmological mean density. Other authors use in their work different criteria to select the linking length parameters. Berlind et al. (2006,[3]) define the linking lengths as:

$$D_L = b_\perp \bar{n_g}^{1/3} \qquad V_L = b_\parallel \bar{n_g}^{1/3} \tag{3.9}$$

where $\bar{n_g}$ is the mean number density of galaxies, $b_\perp$ and $b_\parallel$ are the projected and line of sight linking lengths in units of the mean intergalaxy separation, respectively. In geometrical terms, the resulting linking volume around each galaxy is very similar to a cylinder, oriented along the line of sight, width equal to the projected linking length and whose height is equal to twice the line of sight linking length. It is not a perfect cylinder because its radius increases with redshift, making it slight wider at the far end with respected to the near one, and its base is slightly curved. However, for small linking lengths a cylinder is a good approximation.

[1] The number of galaxies per Mpc per magnitude interval.

3.2 Why Are We Looking for a New Algorithm?

The main feature of the FOF algorithm is that it does not assume or enforce any particular geometry for groups (e.g., spherical), but identifies structures that are approximately enclosed by isodensity surface whose density is monotonically related to the linking lengths. Finally, FOF satisfies a nesting condition: all the members of a group identified with one set the linking lengths are also members of the same group identified using larger linking lengths. The main difficulty is the choice of linking criteria, because the FOF can easily mistake for groups filaments or huge chains of galaxies. Furthermore, the FOF version by Huchra & Geller was designed for spectroscopic data set and due to the relatively large errors in the photometric redshift, it cannot be used for this type of data set without some in deep modifications.
If the photometric errors are not taken into account, then a large fraction of the resulting groups and clusters turn out not to be physically bound objects. The photometric redshift errors are normally larger by a factor 50 with respect to the linking length of the original FOF.
On the other hand, galaxies, with true position on the velocity axis which can be deviant by about $100V_L$, could be combined into groups.
Including the redshift errors in the linking criteria does not solve the mentioned issue, because the resulting structures become unreasonably extended in redshift. Moreover, if large value of linking lengths are used, the FOF find huge chains of galaxies.
For this reason, a new version of this algorithm becomes necessary for photometric redshift dataset. The new algorithm must have the same attractive features of the old one and must take into account the redshift error for each galaxy.
In next section, a new version of FOF based on the idea of Botzler & al. (2004) is presented.

3.3 The NEW FOF

According to Botzler & al. (2004, [4]), the modified version of the FOF algorithm combines the same information as the original FOF, such as Right Ascensions and Declinations, with the use of the photometric redshifts and

takes into account the individual redshift errors of each galaxiy.
The algorithm consists of three parts (see 3.2). In the inner loop, the technique is almost identical to the Huchra & Geller method, with slightly altered linking conditions. The search for groups is carried out in a priori defined redshift slices and only galaxies that are compatible with a given value of the redshift are taken into account. This redshift slicing constitutes the outer loop. So, there is a catalogue of structures for every slice. In a given slice, a galaxy can be the member of only one structure. However, it is important to notice that a galaxy can belong to other structures in other slices.
The third part of the algorithm is the unification across adjacent slices of all structures that have at least a member in common. The *a priori* redshift slices z_{ini} approximates roughly the mean redshift $\bar{V}$ of the original FOF technique. So equation 3.2 and 3.4 change into

$$D_{ij} = 2\sin\left(\frac{\theta_{ij}}{2}\right) D(z_{ini}) \leq D_L \tag{3.10}$$

Where $D(z_{ini})$ is the distance to z_{ini} in Mpc, or

$$D(z_{ini}) = \frac{cz_{ini}}{H_0}$$

The line of sight linking condition is translates into two equations for each galaxy

$$V_i = |v_i - cz_{ini}| \leq \frac{V_L}{2} \tag{3.11}$$

$$V_j = |v_j - cz_{ini}| \leq \frac{V_L}{2} \tag{3.12}$$

If the individual redshift errors δzi and δzj are taken into account, the 6.3 change in

$$V_i \leq [(V_L/2^2 + (c\delta z_i)^2]^{1/2} \leq \frac{V_L}{2} \tag{3.13}$$

$$V_j \leq [(V_L/2^2 + (c\delta z_j)^2]^{1/2} \leq \frac{V_L}{2} \tag{3.14}$$

The first step for this algorithm consists in the choice of the minimal redshift $z_{ini} = z_{min}$. Hence a catalogue of groups that belong to this z_{ini} -slice, is created as it follows.
An object i belonging to this redshift slice that has not yet been assigned

to any group, is chosen from the catalogue.
This galaxy has to be compatible with the chosen z_{ini}. The galaxy j, candidate to be a i friends, must satisfy equation 6.3 and must be closer to i than the cutoff distance D_L. If no friends are found, object i is moved to a list of isolated objectes (like original FOF). If the friends are found, objects i and j are added in the k-list. Finally, a group candidate is called a "real group" if the number of galaxies into the group is:

$$N_{k(z_{ini})} \geq N_{min}$$

The group is then added to the catalogue of structures for the given redshift slice. If it not is case, the galaxy i and its friends are moved to the list of isolated objects. The algorithm goes ahead with the next galaxy from the catalog and its surrounding are searched for friends.
When all galaxies into the $z - slice$ have been processed, the search routine continues in the next redshift slice.
The z_{ini} is increased by a value Δz_{ini} and the original catalog of galaxies is restored. The search for friends is then repeated as described above.
The algorithm produces $(z_{max} - z_{min})/\Delta z_{ini}$ catalogs of clusters and groups, where z_max and z_min are the minimum and maximum redshift of the galaxies into input catalog. The loop ends when $z_{ini} = z_{max}$. When all $z - slice$ are been considered, the unification of structures starts.
This new algorithm, like Huchra & Geller's version, is commutative and yields reproducible result.

3.4 Purpose of this thesis

In order to derive parameter values (such as link lengths, the number of slices, etc.) that are as accurate as possible, the two algorithms are tested on simulated catalogs. These catalogs are derived through the use of software within the GODFinGER project, that is fully described in Chapter 4. Two catalogs are generated: the first provides a simulation of the universe, the galaxies which have spectroscopic redshift, the second one generates the same portion of the universe (thus the same number of galaxies and the same number of simulated teams), but the galaxies which have photometric redshift with an error $\sim \pm 0.02$.

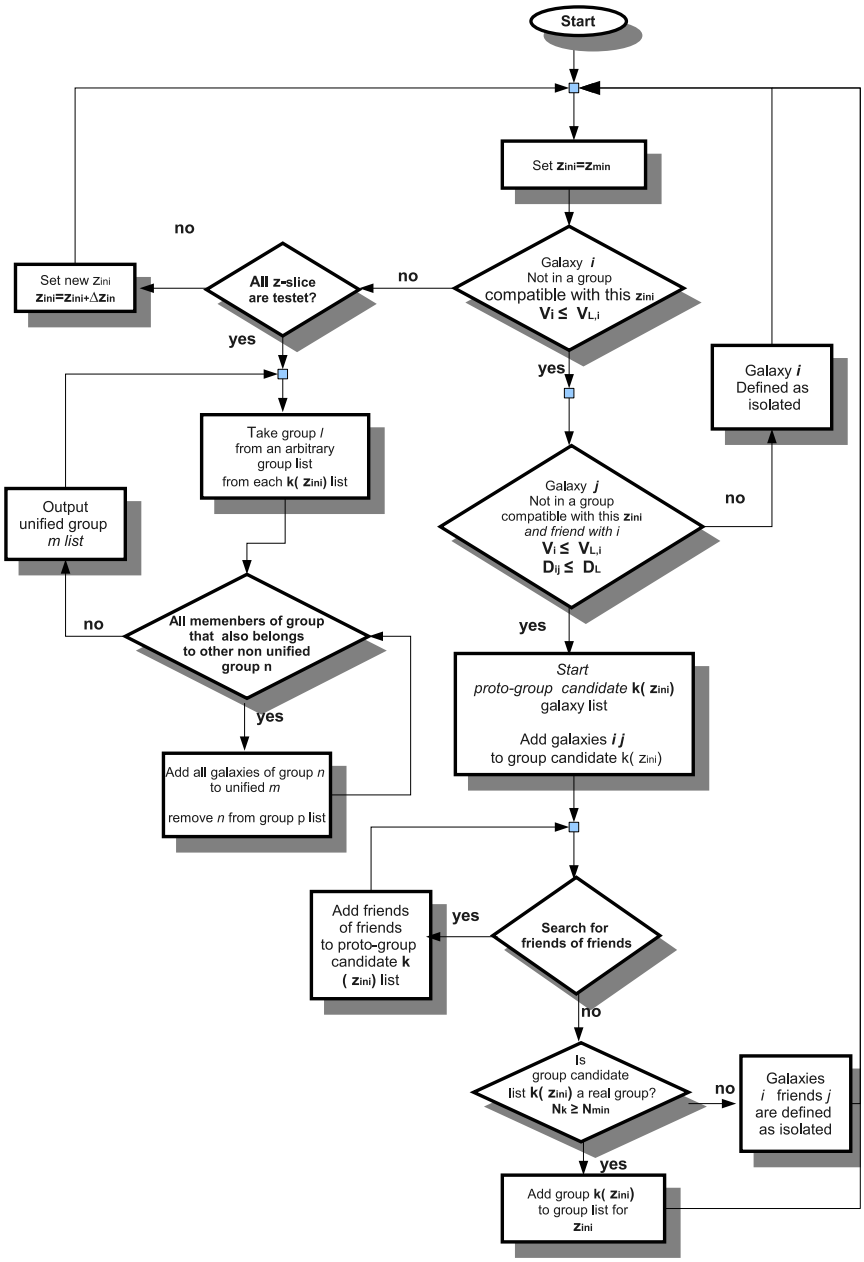


Figure 3.2: Flow chart for the NEW FOF algorithm

As mentioned in the section 3.1.1, the fundamental step is the choice of the correct values for the linking lengths.
The right combination of linking lengths is calculated studying the results obtained from the output of the original FOF run on the mock catalogues for a grid of values of linking lengths.
The right choice of the slicing step for the NEW FOF is calculated running this algorithm on the spectroscopic simulated sample.
Once the linking lengths and the number of slices that must be created, the NEW FOF is applied to the photometric sample generated from the simulation.
The result of this test are discussed in Chapter 5. Finally, the two algorithms are applied to the real data extracted from the Sloan Digital Sky Survey (see, chapter 6)

Chapter 4

GODFinGER

The GODFinGER (*Galaxy Objects Detection Finalized to Groups Extensive Recognition*) project plays the role to fulfill the requested availability of a simulated universe model, populated by groups of galaxies, having a simple structure close as much as possible to the reality. It is worth to highlight that its main purpose consists of the design and implementation of a set of data mining/exploring tools, working on the simulations, aimed at the characterization of a portion of realistic universe. So far, the principle is to create a simulated universe based on real data and cosmological constraints. The real data are coming from those already published by the SDSS.
Usually a mock catalogue of galaxies groups is created by a simulation of universe which requires size, internal structure and the fulfillment of the principal physical and cosmological principles.
Main difference between the GODFinGER methodology and other types of simulations, is that in this case, it is not necessary to simulate the internal dynamics of the Universe, based on the gravitational pull between bodies and their dynamic evolution, because the ultimate aim of the project is the synthesis of a static catalog of groups in place useful to validate reconstruction algorithmssuch as the 3D FoF.
Therefore, instead of complex N-body simulations, the algorithm is designed to model a realistic 3D Universe populated by selected groups of galaxies and background. By Identifying them with point objects, respecting the main constraint coming from the spatial distribution of internal mass, each group is assigned on the basis of their richness. Next sections describe the procedure as it is designed and implemented starting from the scientific space

parameters, proceeding with the software architecture and concluding with preliminary results.

4.1 The Scientific Purpose

The primary purpose of the model is to provide a 3D simulation of a defined portion of the universe, populated by groups and clusters of galaxies, with variable richness, hosted into a cosmic background. So far the basic elements are galaxies, represented by objects or approximated points (mass unit). It is important to clarify that for this methodology, the choice of topology should be made in such a way that the groups or clusters can be randomly positioned but maintaining still identifiable their composite structure, i.e. being able to distinguish between over-density objects (galactic structures in spherical symmetry) and pure cosmic background. In other words the number of groups of a given richness and their 3D structure must comply with specific topological and physical requirements, including the size of the universe limited to the considered volume, the inter-distances between the member galaxies, the variety and range as a function of richness. Finally, it is not possible to neglect that the resulted mock catalogue must be compliant with the physical constants of the cosmological model used. From the mock catalogue it is expected that the size related to object depth (redshift coordinate) will apply an approximately realistic distortion ($z = 0.02$, corresponding to about $6000 km/sec$). Hence, in this way, the mock data set can be considered quite realistic, in comparison with the data obtained from the Sloan Digital Sky Survey.
In order to implement the simulated universe and to populate it with groups and clusters of galaxies, MATLAB results to be the best and flexible modeling environment. The detailed flow-chart for the designed algorithm is shown in 4.1. It reports all basic steps, starting from the parameters definition up to output files creation containing the created model. The simulation results are summarized in the table 3.1 4.1.

4.2 Simulation Model

In order to match the astrophysical requirements, the simulated model requires the definition of some input parameters, such as for instance the

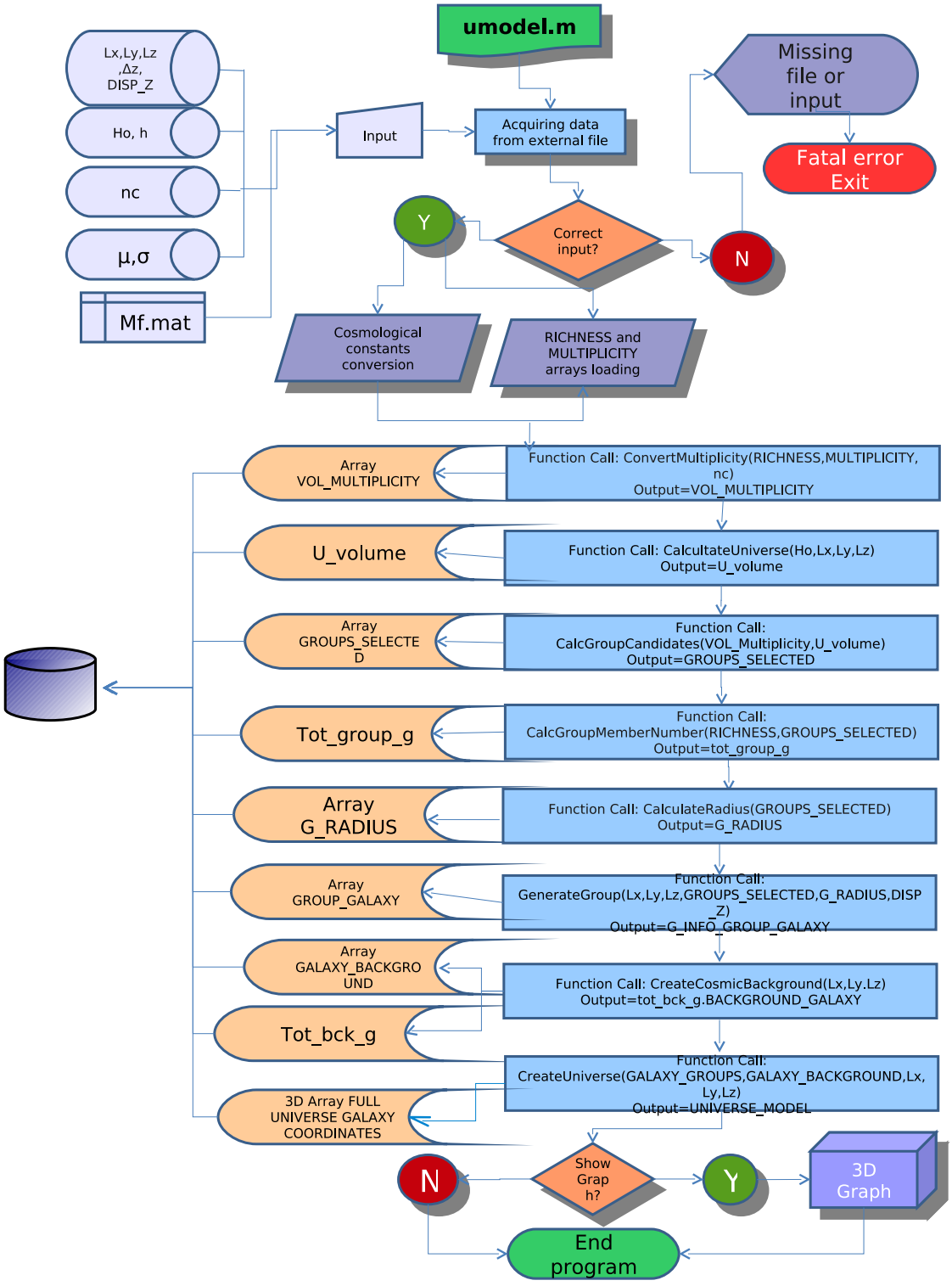


Figure 4.1: **Flow chart of GODfiGER PROJECT.**

Groups	Galaxies	Background Galaxies	Volume
470	2495	5408	$4.22 \times 10^5 Mpc$

Table 4.1: **Features of mock catalogue**.
In this table the features of the mock catalogue are summarized: the number of generated groups ,with the number of galaxies that form both groups and the cosmic background in the mock universe. The last column shows the volume of the mock universe. The size of three dimensions are: $Lx = 20Mpc$, $Ly = 20Mpc$, $Lz = 1056.3Mpc$. It is important to observe that for the third coordinate (redshift) the limit must be 0.25, with a step compliant with phototometric redshift error (~ 0.02) resolution.

size of the universe to be created ($L_(x)$, $L_(y)$, $L_(z)$), obtaining the universe volume, the value of Hubble constant used ($H_0 = 72Mpc^{-1}kms^{-1}$) and the cosmological constants ($\Omega_M = 0.3$,$\Omega_L = 0.7$).
Another fundamental parameter needed to set up the mock catalogue is the value of multiplicity function. For this purpose the values for the multiplicity functions are taken in according to Bachall table (1996). However, this table uses Bachall specific cosmological constants, while for this algorithm has been chosen the (H_0,h) as defined above. Thus, it is necessary a further step to convert these values according to cosmological model in use. Moreover, based on the volume and the multiplicity per unit of volume under consideration, the number of possible groups is estimated according to the richness that can be inserted into the simulated portion of Universe. Furthermore, based on the number of allowed groups and their wealth, the number of galaxies in a group is then created. In addition, each group is selected, based on the simulated radius that defines the maximum dimension limits for both x and y coordinates, defining the number of galaxies allowed to belong to the group. This number is calculated and assigned to each group as follows:

$$\forall \mathcal{R}_i,\ R(\mathcal{R}_i) = \frac{(\mathcal{R}_i - 2)^{0.3}}{4} \tag{4.1}$$

It is now possible, to randomly generate the list of 3D coordinates of galaxy members for each group.This list is generated following both uniform and Gaussian distributions. First of all, the centers are calculated in the three coordinates and then the galaxies are distributed within the range considered. In practice, given the center of the group, its members are distributed in a sphere of radius R (see fig.4.2).

Finally, a randomly population of galaxies is also generated (see fig.4.3).

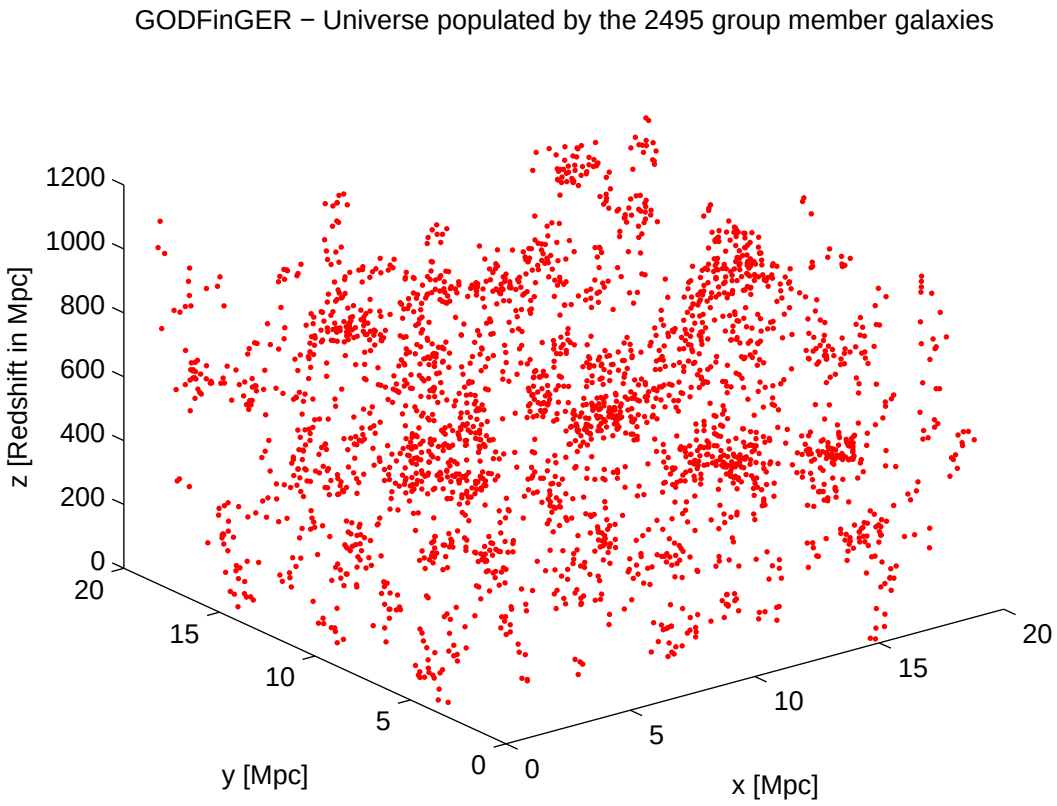


Figure 4.2: **The galaxies members of groups**.

This additional population represents the galaxies of the cosmic background for the mock universe.
The output of the simulation is a set of galaxy coordinates (x, y, z [Mpc]), together with radius and the multiplicity of each group.

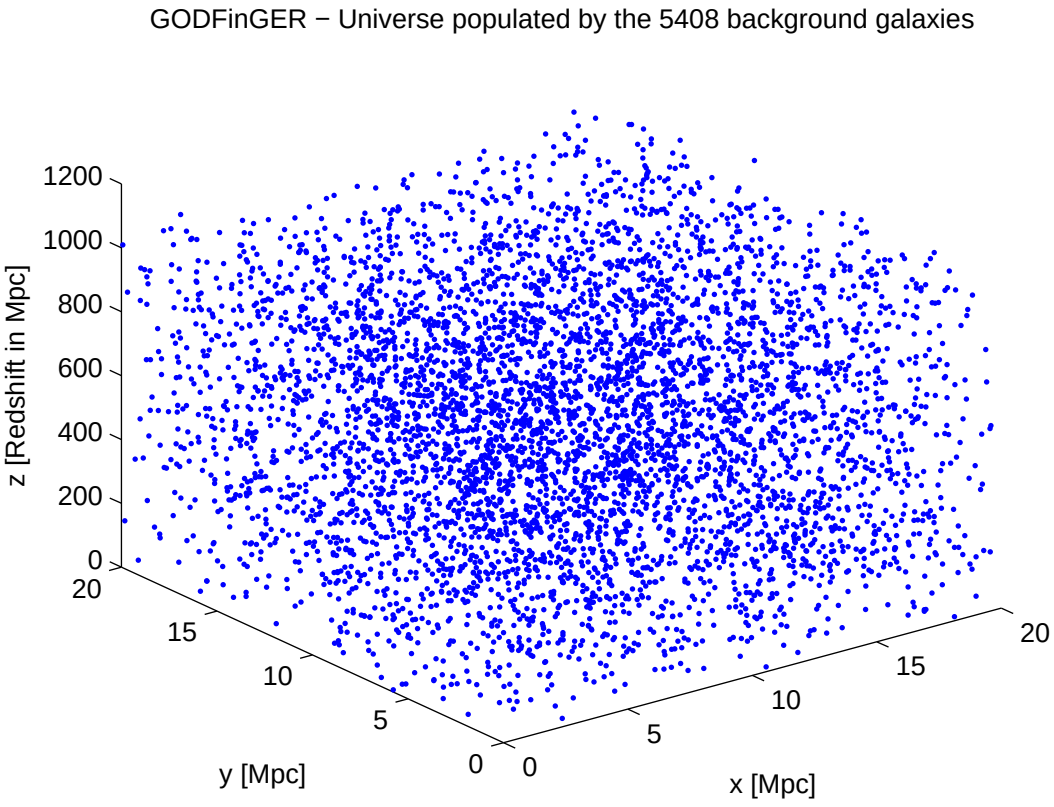


Figure 4.3: The bakground mock galaxies

.

Chapter 5

Testing the FOF algorithm

Properties of groups and clusters of galaxies depend on the algorithm used for their selection. Groups are difficult to find in galaxy catalogs, because the selection suffers from projections and contamination from fore/back ground galaxies. Most previous identifications of group members have been based either on limited or subjective data, or on two-dimensional criteria.
The members of the de Vaucouleurs (de Vaucouleurs, 1975) groups, for example, were identified on the basis of similarity in redshift, apparent magnitude, and morphology as well as on positional coincidence. This method suffers from poorly defined sampling and selection criteria.
The two-dimensional method of Turner and Gott [15] identified group members on the basis of regions were number counts are enhanced with respect to the surrounding environment. This technique suffers from many biases among which the fact that the projected spatial separations vary with the distance and nearby groups projected on large angular scales become so diluted that they cannot be identified. This technique, in fact is usually applied to different magnitude -limited sample different distances. Moreover, this method will not lead to the selection of the same groups when applied to samples which cover the same region of the sky but have different limiting magnitudes.
Moreover, the ongoing quest for a much more robust detection algorithm found more motivation with the necessity of analyzing wide areas of the sky [10].

5.1 The choice of Linking Lengths

It is important to remark that the original FOF algorithm can be used only for data set of galaxies which have spectroscopic redshift, to finalize the algorithm we built a mock catalogue containing 7903 galaxies distributed among 470 groups and an uniform background. As mentioned in sec. 3.1.1, the choice of linking lengths is the most important step. In order to find the right combination of linking lengths in fact, results obtained from the output of the original FOF fed with the mock catalogue for a grid of linking lengths are evaluated. Particularly, two are the features of the obtained group distribution that were evaluated: the first one is the relationship between the number of galaxies within the group which are present in the simulation catalogue, N_{true} *true groups*, and the galaxies number,N_{obs}, present in the group catalogue calculated by the FOF algorithm. The second one gives the multiplicity function.
Since the original FOF can be used only for data set of galaxies which have spectroscopic redshift, the mock catalogue was generated with this kind of redshift. The two linking lengths are defined (see 5.1) according to Berlind et al. ([3]), as

$$D_L = b_\perp \bar{n_g}^{1/3} \qquad V_L = b_\parallel \bar{n_g}^{1/3} \tag{5.1}$$

where $\bar{n_g}$ is the mean number density of galaxies; while the two parameters $b_\perp$ and$b_\parallel$ are free to vary within the set of linking length values generated as follows:

$$0.14 \leq b_\perp \leq 0.2 \qquad 0.5 \leq b_\parallel \leq 2.0$$

Usually in original FOF algorithm, the ratio of radial and transversal linking lengths $b_\parallel/b_\perp$ is taken as constant. In according with the choice of these parameters made by Berlind et al. [3], wider ranges were excluded from ours analysis. So, considering the value of $b_\parallel$ and $b_\perp$ described above, the value of this ratio are included between

$$1.3 \leq b_\parallel/b_\perp \leq 14.29$$

. For each set of parameters, resulting catalogue groups are analyzed and compared with the simulated one. Particularly are compared both galaxies in group and multiplicity function for both simulated catalogue and derived

catalogue.

5.1.1 N_{true} vs N_{obs}

The group multiplicity function is an average statistical property showing the abundance of structures as a function of richness. Therefore, at least in principle, it is possible that the resulting MFs are identical without the need for the individual groups being identical. For the mentioned reason, before deriving the multiplicity functions for each set, an unbiased relation between the multiplicity of individual groups, and N_{true} is required.
This was done performing a match between simulated and recovered groups of galaxies on a one-to-one basis. In this way, it is possible to estimate the number of galaxies present in the simulation but not in FOF catalogue and vice versa. Additionally, it is possible to evaluate the differences between the two distributions.
The linking length ratio was chosen to minimize the differences between two samples:

- the *spurious* sample: the FOF group (or galaxies) that are not associated with any simulated group;
- *undected* sample: the groups (or galaxies) that are not found by FOF;

Results are listed in table 5.1. In this table there are result obtained by running the FOF algorithm on the simulated data set, varying the linking lengths. In tables 5.2 and 5.3, are completeness and contamination for each pair of parameters $b_{\parallel}$ and $b_{\perp}$.

Furthermore, the comparison between mock and FOF catalogues allows to understand how the galaxies are distributed into FOF groups. Most of undetected galaxies are associated by FOF into groups that have one or two galaxies more than the corresponding mock group.
It is important to observe that it is possible that FOF groups are a combination of mock groups but also some mock structures can be split into more FOF groups. This depends on the choice of linking parameter. Increasing the value of linking lengths, in fact, the rate of these groups increases. The best choice turns out to be $b_{\parallel}/b_{\perp} = 2.5$. In this catalogue, only one group is the combination of two mock ones. Similarly, only one mock group is split into two groups by FOF.

$b_{\parallel}/b_{\perp}$	**FOF Groups**	N_{gal}	**Number Spurious Group**	**Number Spurious Galaxies**	**Number Undetected Groups**	**Number Undetected Galaxies**
14,29	470	2479	226	1285	224	1301
10,71	464	2429	17	142	25	208
8,58	487	2704	32	252	5	45
7.14	453	2366	2	46	28	215
5,36	453	2343	4	67	28	219
5,00	361	1702	0	21	134	814
4,29	468	2561	9	112	6	46
4,17	453	2366	0	33	25	208
3,57	452	2325	20	52	29	222
2,86	466	2514	2	67	6	48
2,50	466	2562	2	80	2	13
1,33	602	6582	153	4087	0	0

Table 5.1: **Comparison of the FOF (and mock) catalogue** In the col. 1 there are different linking lengths ratio used; in col. 2 and 3 there are the groups found by FOF and the galaxies in these groups, respectively. Spurious galaxies and groups are shown in col.4 and 5; in col. 6 and 7 are listed "Undetected" groups and galaxies.

$b_{\perp} \rightarrow$	0.10	0.12	0.14	0.17	0.20
0.50	134	25	29	6	2
0.75	121	67	28	6	2
1.00	118	66	28	6	2
1.50	116	64	25	5	210
2.00	112	59	224	4	2
$b_{\parallel} \uparrow$					

Table 5.2: **Undetected groups for each linking lenghts pairs**
This table shows how the number of spurious groups changes for each linking lengths pair.

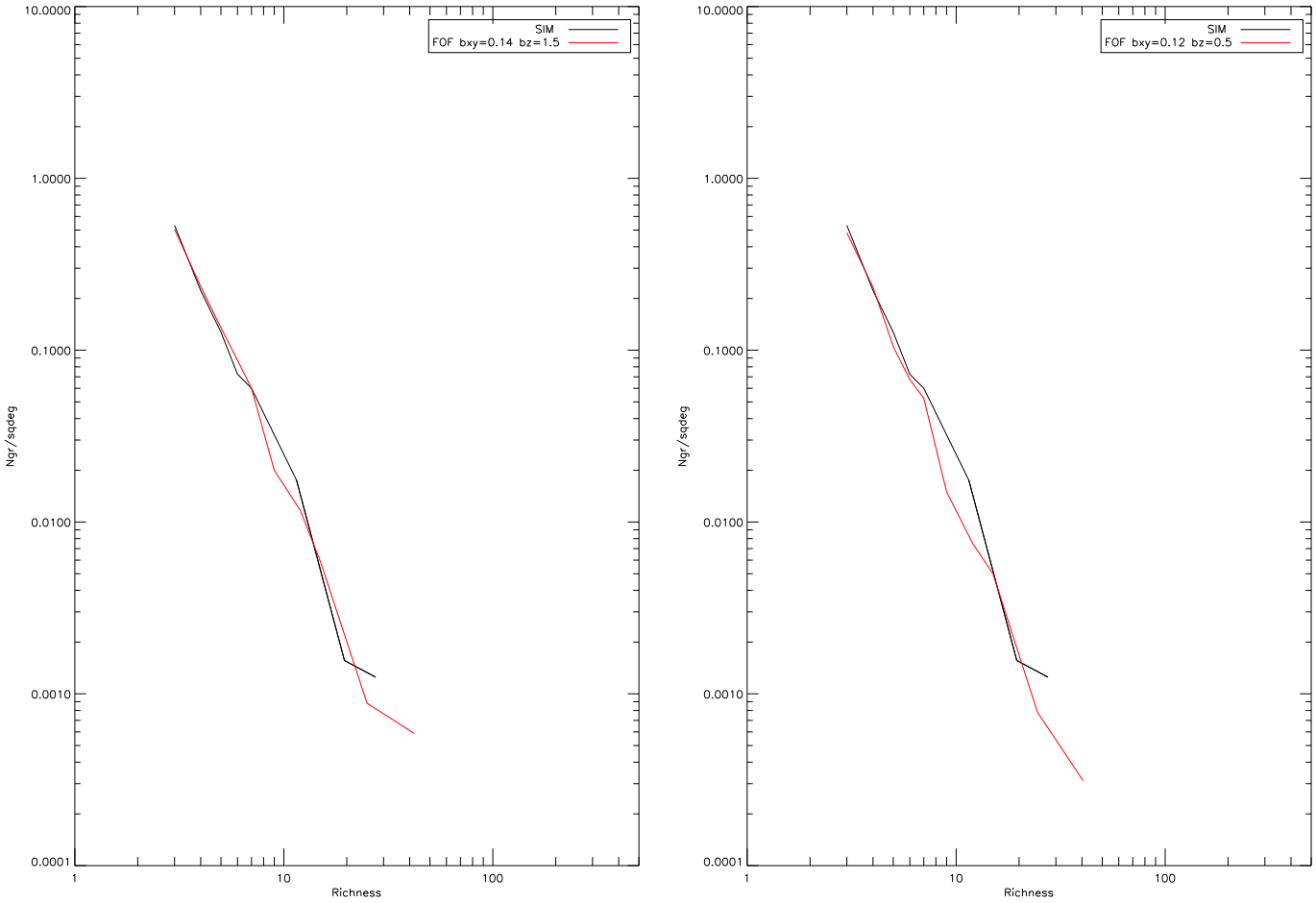
$b_\perp \rightarrow$	0.10	0.12	0.14	0.17	0.20
0.50	0	0	20	2	2
0.75	0	1	4	9	13
1.00	1	2	2	18	29
1.50	3	6	17	32	258
2.00	4	9	226	47	2
$b_\parallel \uparrow$					

Table 5.3: **Spurious groups number for each linking lengths pairs**
This table shows how to change the number of spurious groups for each linking lengths pair.

5.2 Comparison of Multiplicity Function

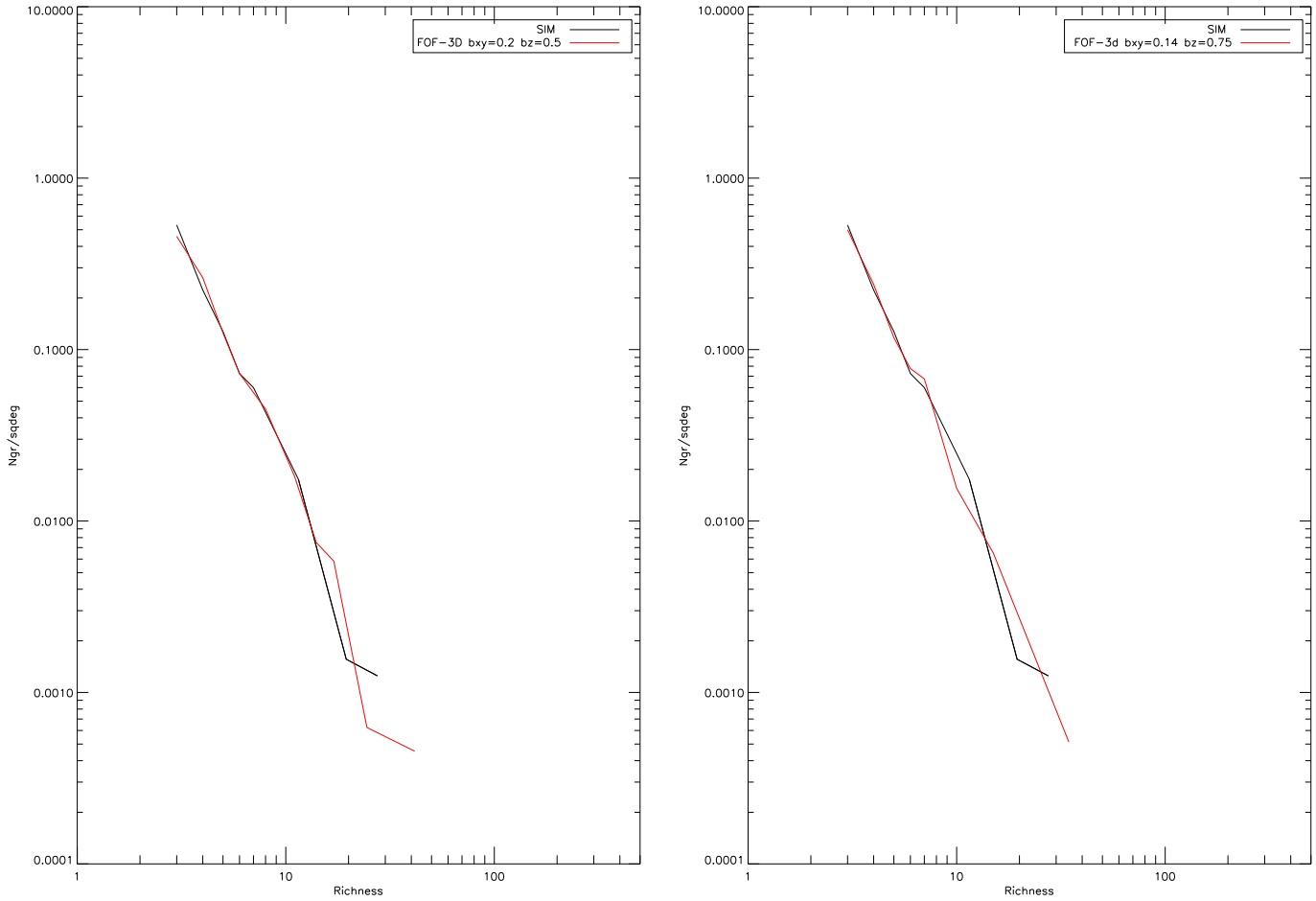
Another feature analyzed for each catalogue is the MF. Such comparison results more easily when both MFs are plotted in the same chart. In fact in figure 5.1 and 5.2 the MF for the mock catalogue and for the one resulted from the application of FOF algorithm are plotted in black and in red, respectively.
Analyzing more in details the graphs, particularly panel a) in Fig. 5.1, it is possible to observe that this curve reflects the good performance of MF for low-richness ($3 \leq \mathcal{R} \leq 10$) groups; however it is underestimated the number of intermediate richness groups. Panel b) in the same figure it is obtained making use of the parameters $b_\perp = 0.12$ $b_\parallel = 0.5$. The graph clearly show that these parameters underestimate the linking lengths, hence the number of groups found by FOF is lower when compared with the simulation groups number (453 compared with 470). In fig 5.2 panel a), the MF of the catalogue obtained with $b_\perp = 0.14$ $b_\parallel = 1.5$ is then compared with simulation one. As it is possible to infer from the chart, these are the best values, in fact, the two curves are similar for groups of low and average richness. In panel b) it is compared the MF obtained making use of the parameters value: $b_\perp = 0.14$ $b_\parallel = 0.75$, (that are the ones used by Berlind et al [3]), with the simulation MF. The graph clearly shows that the MF of this catalogue differs very much from the simulation, especially for groups at high multiplicity.

Objective comparison analysis between the mock MF and the FOF one is also performed through a Kolmogorov-Smirnov statistical test. The results of this test are shown in tab 5.4. As it is possible to read from the table, the best value for the linking lengths ratio turns out to be 2.5. In fact according



(a) (b)

Figure 5.1: **Examples of Multiplicity Function Original FOF.**
In this figure, the multiplicity functions derived from Original FOF with different linking lenghts are shown. a)$b_\perp = 0.14$ $b_\parallel = 1.5$. b)$b_\perp = 0.12$ $b_\parallel = 0.5$.



(a) (b)

Figure 5.2: **Multiplicity Function Original FOF**.
In this figure, the multiplicity functions derived from Original FOF with are shown.
a) $b_\perp = 0.2$ $b_\parallel = 0.5$; b) $b_\perp = 0.14$ $b_\parallel = 0.75$.

$b_{\parallel}/b_{\perp}$	**P**	**D**
14.29	0.070	0.44
10.71	0.087	0.43
8.57	0.11	0.41
7.5	0.09	0.42
7.14	0.07	0.44
5.36	0.08	0.43
5.0	0.024	0.52
4.29	0.05	0.46
4.17	0.14	0.40
3.57	0.11	0.42
2.86	0.07	0.44
2.5	0.17	0.38
1.3	0.02	0.46

Table 5.4: **Result of Kolmogorov-Sminorv Test**. In the column 1 there are the linking lengths ratio used to generate the catalogue. In the column 2 there are the probabilities that two distributions of data values are drawn from the same distribution. The value in the column 3 specify the maximum deviation between the cumulative distribution of the data and the supplied function.

to the tests, the FOF catalogue that exhibits the smallest fraction of spurious groups and the highest completeness is obtained using the linking lengths ratio 2.50 and in particular,

$$b_{\perp} = 0.2 \qquad b_{\parallel} = 0.5.$$

Using these values, the obtained catalogue just has two undetected groups and two spurious groups. Moreover, the richness of the undetected groups is always equal to 3, while all the richer mock groups are found by FOF.

5.3 New FOF runs on simulated data set

The modified FOF algorithm (see Chapter 3), was tested on the mock catalogues and its performances were evaluated against those of the original FOF. For this reason, the New FOF is firstly tested on a the same spectroscopic mock catalogue, used for the original FOF.
It is important to recall that even though the New FOF was specifically built to work on photometric redshifts, it can also be used on spectroscopic

data. In this case, the individual redshift errors δz_i, δz_j are not taken into account, so the linking lengths conditions on the line of sight become,

$$V_i = |v_i - cz_{ini}| \leq \frac{V_L}{2} \tag{5.2}$$

$$V_j = |v_j - cz_{ini}| \leq \frac{V_L}{2}. \tag{5.3}$$

As for the projected linking length, the correct value depends on the surface density of each slice, so that

$$D_L = b_\perp \bar{n_{gs}}^{1/2} \tag{5.4}$$

In this way, D_L varies with redshift, while for the original FOF it was a constant. In this part of the test, the number of slices to be created and the value of incremental step per z-slice, Δz_{ini}, must be chosen as well.
It is evident that the choice of a large Δz_{ini}, led to loose a large number of galaxies, both increasing the possibility that rich structures can be separated and loosing low multiplicity structures. As summarized in Figure 5.3, increasing the steps Δz_{ini}, the number of group reaches a plateau. In order to obtain a number of groups as close as possible to the one present in the simulated data set, and at the same time to improve the computational time, the better methodology consists in to the use of $\Delta z_{ini} = 0.1 Mpc$. Considering that the maximum values of the line of sight distance and the smallest one are

$$D_{z_{max}} = \frac{cz}{H_0} = 1056 Mpc$$

and

$$D_{z_{min}} = \frac{cz}{H_0} = 0.024 Mpc$$

respectively, the corresponding number of slice is

$$(D_{z_{max}} - D_{z_{min}})/\Delta z_i ni = 10564$$

.

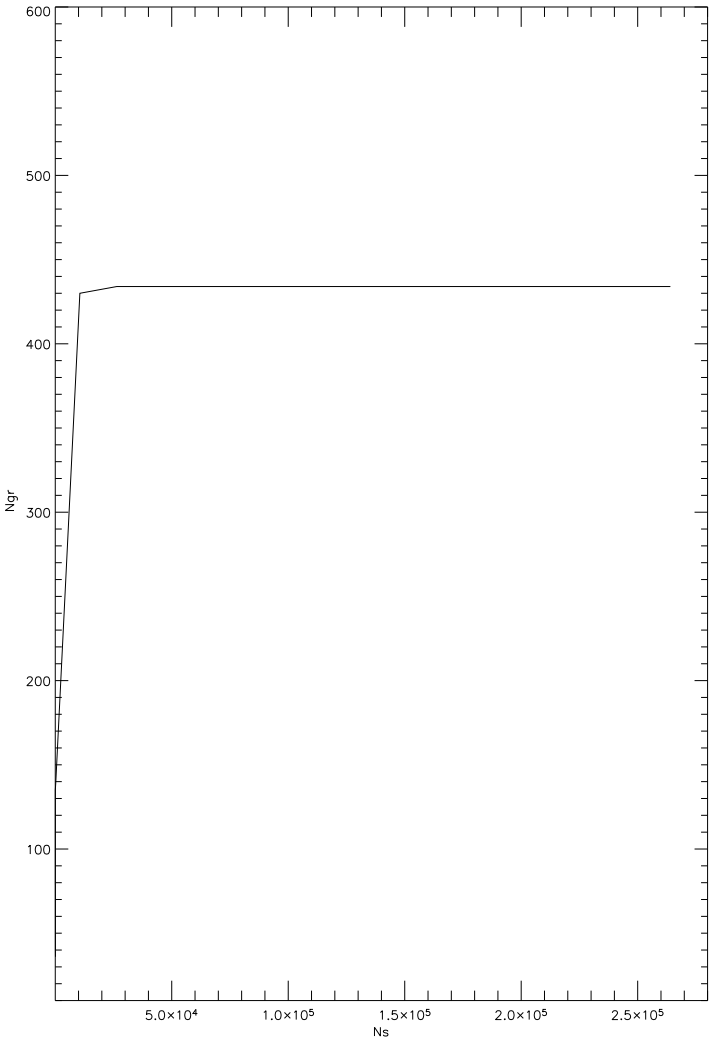


Figure 5.3: **Trend in the number of groups** (N_g) **in function of the number of slices** (N_s) **per New FOF.**
This plot is obtained varying the step between $0.004Mpc$ and $50Mpc$. It is possible to observe, that if the number of slice increases (decreasing the step value) the number of groups in the sample gets to a plateau. These catalogue are obtained using $b_\perp = 0.2$ and $b_\parallel = 0.5$, that is the best value compared with the one fond by the original FOF running on the same catalogue.

5.4 Comparison of group catalogue

A second spectroscopic catalogue of groups was obtained by running the New FOF algorithm on the mock galaxies sample. The linking lengths are the same used for the original FOF algorithm; moreover the slicing step value is $\Delta z_{ini} = 0.1 Mpc$, as already mentioned above.
The resulting catalogue of structures contains 430 groups, for a total number of 1994 galaxies. We then analyzed the differences between this catalogue and the simulated groups catalogue, in order to provide the number of spurious and undected groups. Then, a comparison between the results obtained by the NEW FOF and ones obtained by the original FOF is provided.

5.4.1 New FOF and Simulated Data

The completeness of the catalogue, obtained with the NEW FOF $\sim$ 85%. Although this is a high value, it is low compared to the completeness obtained by the original FOF ($\sim$ 99%).
By comparing the multiplicities function, it is possible to observe, that the linking lengths value seem to be too low (see, fig.5.4). In fact, there is a good match for groups with low richness, but a noticeable fraction of groups with high richness remains undetected. This difference might be related to the fact that the two algorithms differ in both the linking lengths and in the calculation of the distance between two galaxies i and j. In fact, the Orginal FOF algorithms uses a constant projected linking length, while the New FOF uses a linking length which varies with z. In addition, the Original FOF calculates the distance between two galaxies using the relationship

$$D_{\perp_{i,j}} = \frac{c}{H_0}(z_i + z_j)\sin(\frac{\theta_{ij}}{2}) \tag{5.5}$$

where θ_{ij} is the angular separation between i and j, z_i and z_j are the redshift of two galaxies. The New FOF, instead, uses

$$D_{\perp_{i,j}} = \frac{c}{H_0} z_{slice}\sin(\frac{\theta_{ij}}{2}) \tag{5.6}$$

where z_{slice} is the redshift corresponding to particular slice analyzed.
Taking into account this observation, a catalogue was derived using the New

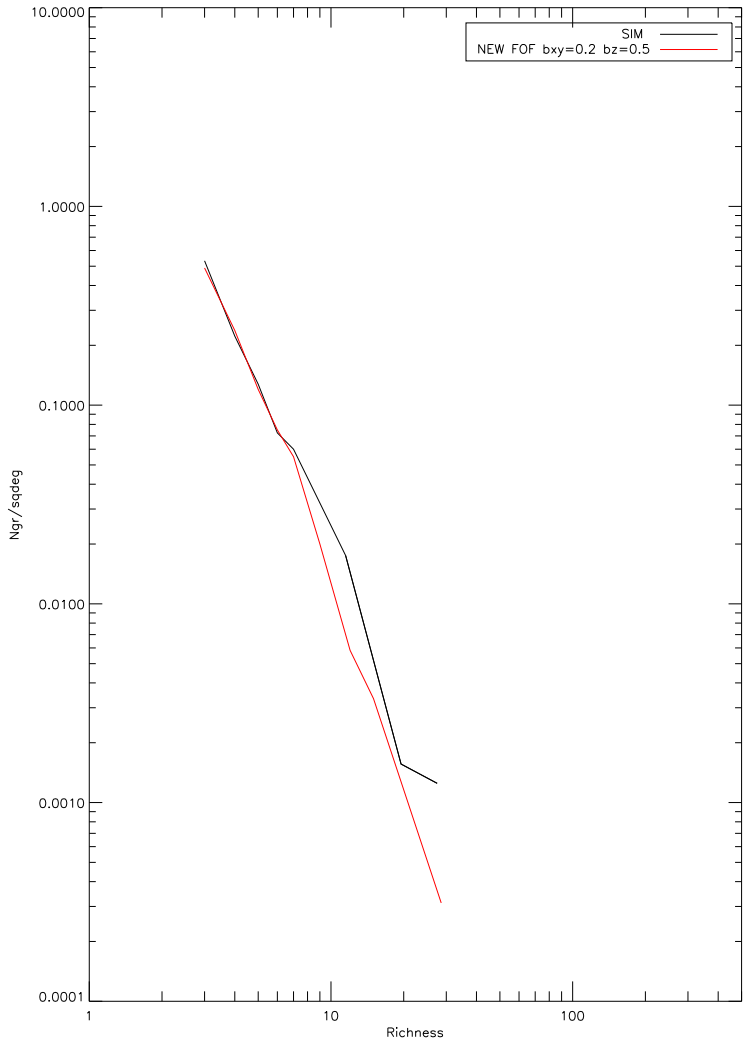


Figure 5.4: Multiplicity Function New FOF. This is multiplicity function (*red line*)for the catalogue obtained by New FOF compared with multiplicity function of simulation catalogue of groups(*black line*). The linking lengths used by New FOF are $b_{\perp} = 0.2$ and $b_{\parallel} = 0,5$.

FOF algorithm leaving unchanged the ratio $b_{\parallel}/b_{\perp} = 2.5$, but with value for the $b-$parameters scoring: $b_{\parallel} = 1$ and $b_{\perp} = 0.4$. With these parameters value, the New FOF algorithm finds 494 (about 64 groups more than in the previous catalogue) groups, comprising 2761 galaxies which 39 groups resulted to be spurious (comprising 289 spurious galaxies) and 2 groups are *undetected* (with 20 undetected galaxies). With this new values the catalogue completeness reaches $\sim (99)\%$.
Unfortunately, the rate of spurious galaxies and spurious groups increases with respect to the catalogue obtained by original FOF. In the previous catalogue, the contamination of spurious groups was $\sim (0.43)\%$, with $\sim (3)\%$ of galaxies in group that are spurious; in the NEW catalogue, the fraction of spurious groups is $\sim (7.89)\%$ (and out of 2761 galaxies in groups, 10% are spurious).
Nevertheless, it is possible to observe that the multiplicity function obtained from this catalogue and from the simulated one, share the same trend (see, 5.5). In the next sections, a more detailed comparison between the results of the two algorithms is provided.

5.4.2 New FOF and Original FOF: analysis of discrepancies

In table 5.5 are summarized the results obtained from both the original and the new FOF algorithms running on the same spectroscopic mock data set. A match between the two catalogues has been done on a one to one

	Groups	**Galaxies in groups**
Original FOF	466	2761
New FOF	494	2562
Simulation	470	2495

Table 5.5: **The result obtained from both the original and the new FOF algorithms.** Comparison of the results obtained to running Original FOF and New FOF on the same input data set. The Total Groups column contains the total number of group found and the Total Galaxies column the number of galaxies distributed in structures. The total "true" number and the true number of galaxies are shown too.

basis. In particular, four types of groups have been identified in the New FOF catalogue:

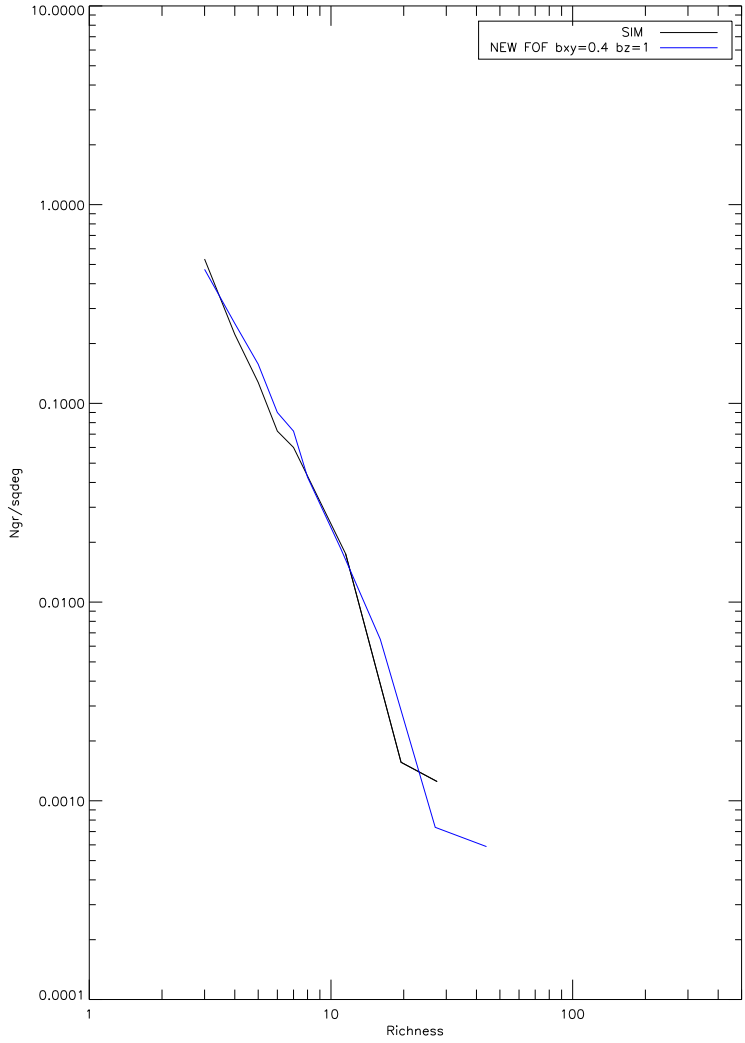


Figure 5.5: **Multiplicity Function of spectroscopic catalogue obtained by running New FOF on the simulation data set.**
The blue line is the multiplicity function obtained running New FOF on the mock catalogue, using $b_{\|} = 1$ and $b_{\perp} = 0.4$ and step $0.1 Mpc$. The black line is the simulation multiplicity function.

1. Structures that are combinations of original FOF groups.

2. Structures that differ by few members.

3. Groups that are found only by the New FOF : no galaxy in a group is a member of the original FOF groups.

4. Identical groups. Groups that have identical composition.

This classification scheme facilitates the examination of the intrinsic characteristic of the two FOF algorithms. Results are shown in the table. 5.6

	New FOF	FOF	%FOF	% New FOF
Only FOF structures	–	2	0.40%	–
Only New FOF structures	39	–	–	8%
Structures different in a few members	66	27	6%	13%
Identical	362	362	77%	73%

Table 5.6: Comparison of the Orginal FOF structure and New FOF ones

In fig 5.6, the multiplicity functions for the original FOF catalogue and the New FOF one are compared.

5.5 New FOF testing on the photometric simulation

The New FOF algorithm is now applied to the photometric data set. Important parameters such as the slicing step, and the number of slices, were evaluated through the previous test, particularly the slicing step is fixed to be equal to $\Delta z_{ini} = 0.1 Mpc$. Moreover, as it was already mentioned, in the photometric data set it is necessary to take into account the photometric redshift errors of each galaxy, therefore for this reason, the linking length criteria in the line of sight direction becomes

$$V_i \leq [(V_L/2)^2 + (c\delta z_i)^2]^{1/2} \leq \frac{V_L}{2} \tag{5.7}$$

where δz_i is the individual redshift errors equal to 0.02 (the 3σ uncertainty on the redshift slice).
The first test performed with the linking lengths parameters found above, $b_{\parallel} = 1$ and $b_{\perp} = 0.4$ gave as result a group catalog containing 149 groups (for a total of 664 galaxies) a maximum of 33. To understand this, the low richness of retrieved groups, the trend of average density on each slice is studied and compared with the distribution of galaxies in the slice, when

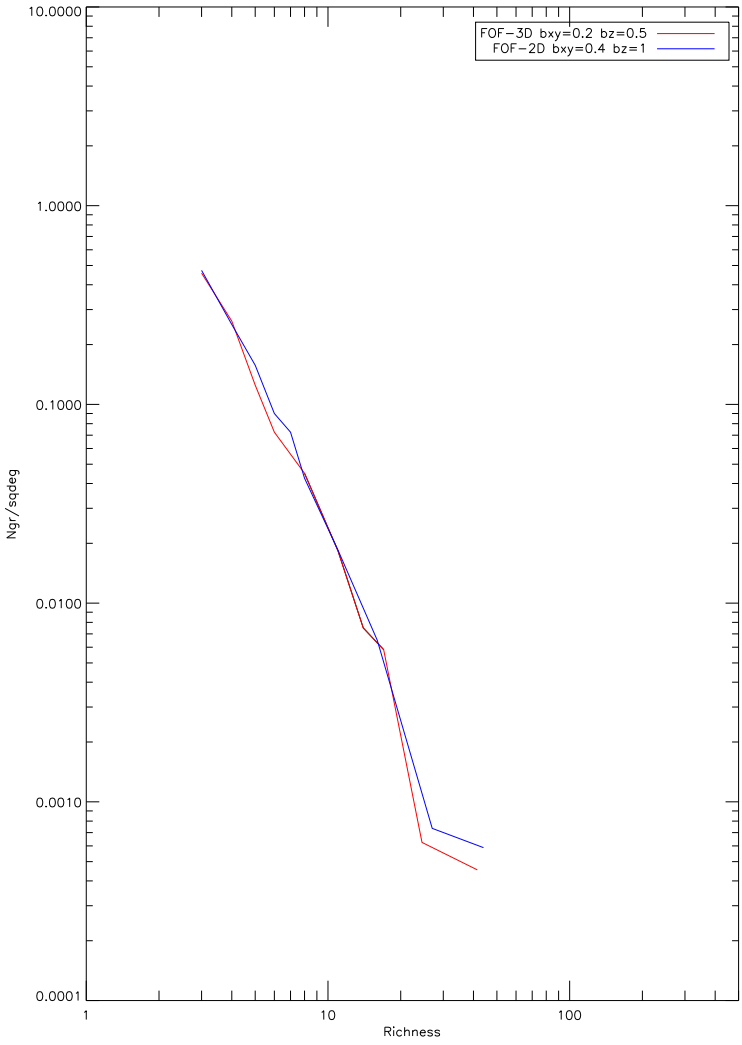


Figure 5.6: Multiplicity Function of spectroscopic catalogue obtained by running New FOF on the simulated data set.
The blue line is the multiplicity function obtained running New FOF on the mock catalogue, using $b_{\parallel} = 1$ and $b_{\perp} = 0.4$ and step $0.1 Mpc$. The red line is the multiplicity function calculated for the spectroscopic data set with the original FOF $b_{\parallel} = 0.5$ and $b_{\perp} = 0.2$

the photometric errors is taken into account. From Fig.5.7, it is possible to observe that when the errors are not considered, or else the $\delta_{z_i} = 0$ into (5.7), the number of galaxies does not change with the $z - slice$. When $\delta_{z_i} \neq 0$, there is a strong dependence on $z - slice$ (see panel b). Therefore, the density of each slice changes with respect to the above mentioned case, and so the projected linking length (recalling that D_L it is defined to be equal $b_\perp \cdot \bar{n}^{-1/2}$. Thus it is necessary to define a new value for $b_\perp$, to correct for this factor.

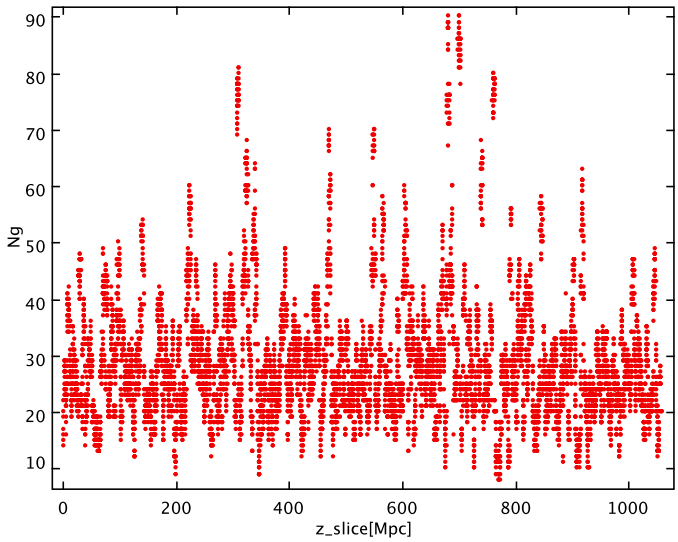
The choice of the new value was been made complying with the criteria of completeness and contamination as described in the previous sections.
The best value resulting from the test was $b_\perp = 0.9$. The resulting catalogue, contained 415 groups, comprising a total of 2960 galaxies. The completeness of this catalogue was 74% and the contamination is 29%.
In order to improve these result, a further test was carried out. The $b_\perp$ value was kept fixed, and other catalogs was generated by choosing the error value 1σ, 0.6σ. Results are shown in 5.7.

Using the errors at $0,6\sigma$ value, the completeness improved (increasing its

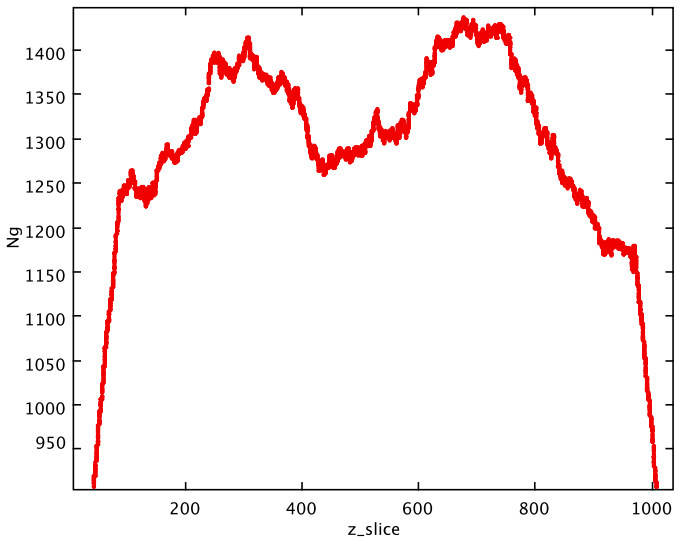
	Total Groups	**Undetected Groups**	**Spurious Groups**
$0,6\sigma$	580	93	154
1σ	479	50	213
3σ	415	122	123

Table 5.7: **The result obtained varying errors**.
Comparison of the results obtained at different σ values.

value from 74% up to 80%). However, also the contamination increased to 8%. Unfortunately it is not possible to vary $b_\perp$ in a manner that would allows to obtain a catalog that results at the same time complete and not contaminated. In this case, comparing the multiplicity functions (see, fig. 5.8), the value for the errors to 3σ resulted to better fit our needs.

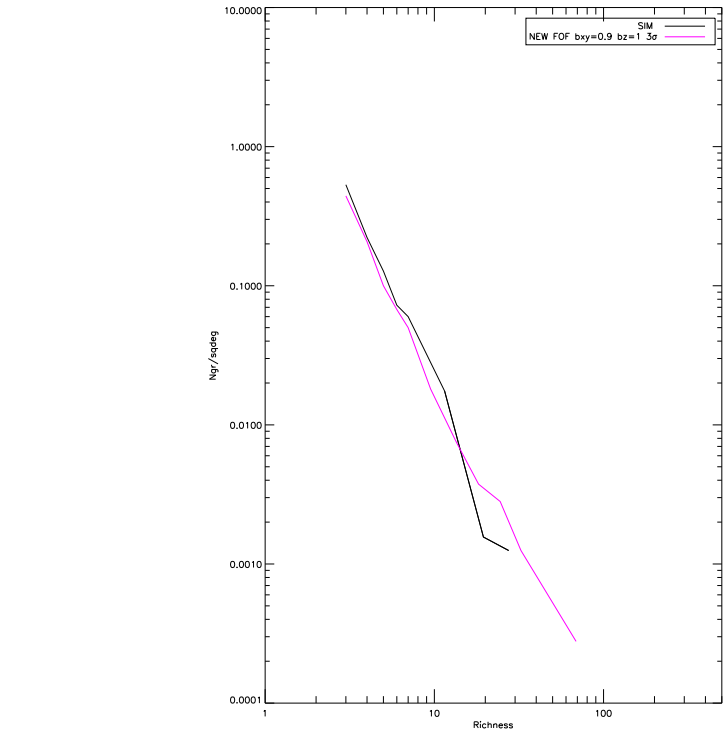


(a)

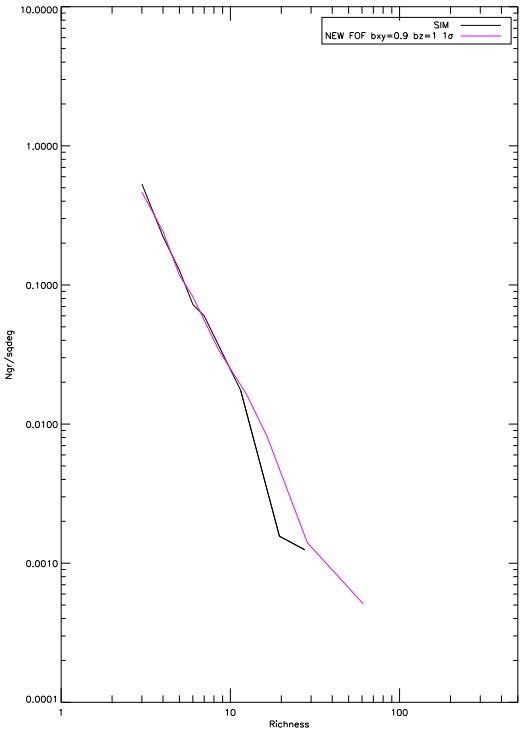


(b)

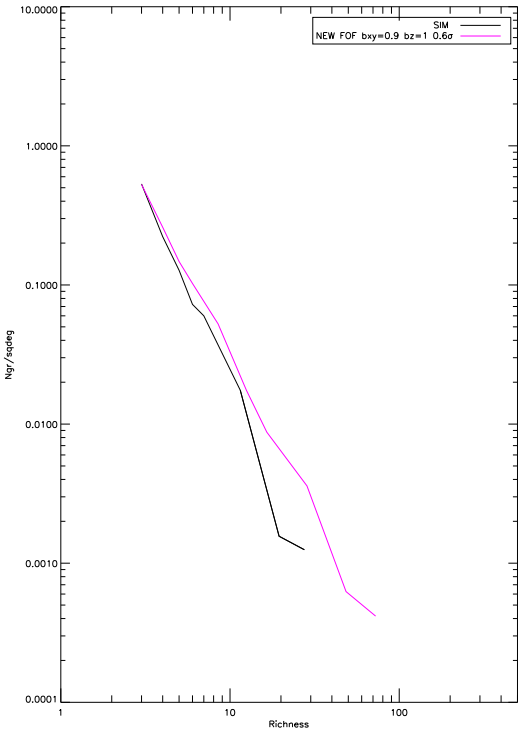
Figure 5.7: **Comparison between the galaxies distribution into a slice.**
Taking into account the individual redshift errors, the galaxies distribution into each slice changes with respect the slices distribution of galaxies when the spectroscopic (without redshift errors) data set is used.
The figure a) shows the trend for the spectroscopic data set. The data used in this plot are obtained running the same algorithms with same linking lengths. The figure b) shows the trend of the galaxies number respect to z-slice, when used photometric redshifts. As it is possible to observe, the number of galaxies in the slice is not constant.



(a)



(b)



(c)

Figure 5.8: **Multiplicity function varying errors.**
The figures a), b), c) are the multiplicity functions calculated 3σ, 1σ,0.6σ

Chapter 6

Application to the Photometric Redshift Catalogue

Before presenting the results obtained from the application of the New FOF on the photometric redshift catalogue, it is worth to clarify an important issue arising from the lack in literature of such kind of catalogue. A broadly accepted way to evaluate the performance of such an algorithm requires the evaluation of the resulting catalogue obtained by the same algorithm turned to run on a photometric sample. Such evaluation usually consists into a performance assessment of the catalogue compared with the one obtained using the original FOF algorithm proposed by Berlind running on his sample. However, it is worth to highlight that in such kind of comparison (and in any case when it is attempted the comparison between two or more catalogues), it is compulsory to pay attention about the type of algorithm used to generate them, about the possible existence of a selection criteria in the data and so on. Failing to comply with the last observation, could result into an imprecise or, even worst, a biased performance evaluation.

Across the following sections, the New FOF algorithm is evaluated following the criteria mentioned above. Firstly, for the sake of clarity, it is necessary to spend a few words about the real dataset used (the Sloan Digital Sky Survey: Fifth Data Release).

6.1 Sloan Digital Sky Survey: Fifth Data Release

The Sloan Digital Sky Survey (hereafter SDSS) is an imaging survey over a contiguous 7646 deg^2 high-latitude elliptical region in the Northern Galactic Cap, plus an additional 750 deg^2 in the Southern Galactic Cap. The SDSS provides a uniform, well-calibrated map in five broad bands (u, g, r, i, z) spanning a wide wavelength range: $3000 - 1000$ Å. The catalog derived from the images includes more than 350 million celestial objects, and spectra of 930.000 galaxies, 120.000 quasars, and 460.000 stars. The data are fully calibrated and reduced, carefully checked for quality, and publically accessible through efficient databases (http:www.sdss.org). The data have been publicly released in a series of annual data releases. The imaging survey is accompanied by almost a million spectra of two complete samples of galaxies and more than 120,000 spectra of quasars, covering about 8200 square degrees. The basic properties of the SDSS are summarized in Table.6.1 6.2. In the Fig.6.1 the distribution on the sky of SDSS imaging (*top*) and the spectroscopy (*bottom*) included in DR5, are shown in $J2000.0$ equatorial coordinates. In this work the Fifth Data Release (hereafter DR5) was used.

In the Fig. 6.1 are shown the regions of the sky in DR5 covered by imaging (*top*) and spectroscopy (*bottom*). In contrast to DR4, the imaging available in DR5 covers an essentially contiguous region of the North Galactic cap, with a few small parches totaling ~ 200 deg^2 still remaining. The area covered by DR5 primary imaging survey is 8000 deg^2.
The area covered by the spectroscopic survey is 5713 deg^2. In Tables 6.1 and 6.2 are summarized the characteristics of the DR5 imaging and spectroscopic surveys, respectively. Note that "stars" and "galaxies" division in tab. 6.1 refer to the photometric pipeline classifications. Classification in 6.2 are those returned by the spectroscopic pipeline.

6.2 Photometric Redshift

Despite the huge effort for a wide spectroscopic coverage, spectroscopic redshift are available only for a small minority of objects. Photometric estimates of the redshift are instead available for all SDSS galaxies. The photometric redshift catalogue used for this work, has been constructed by

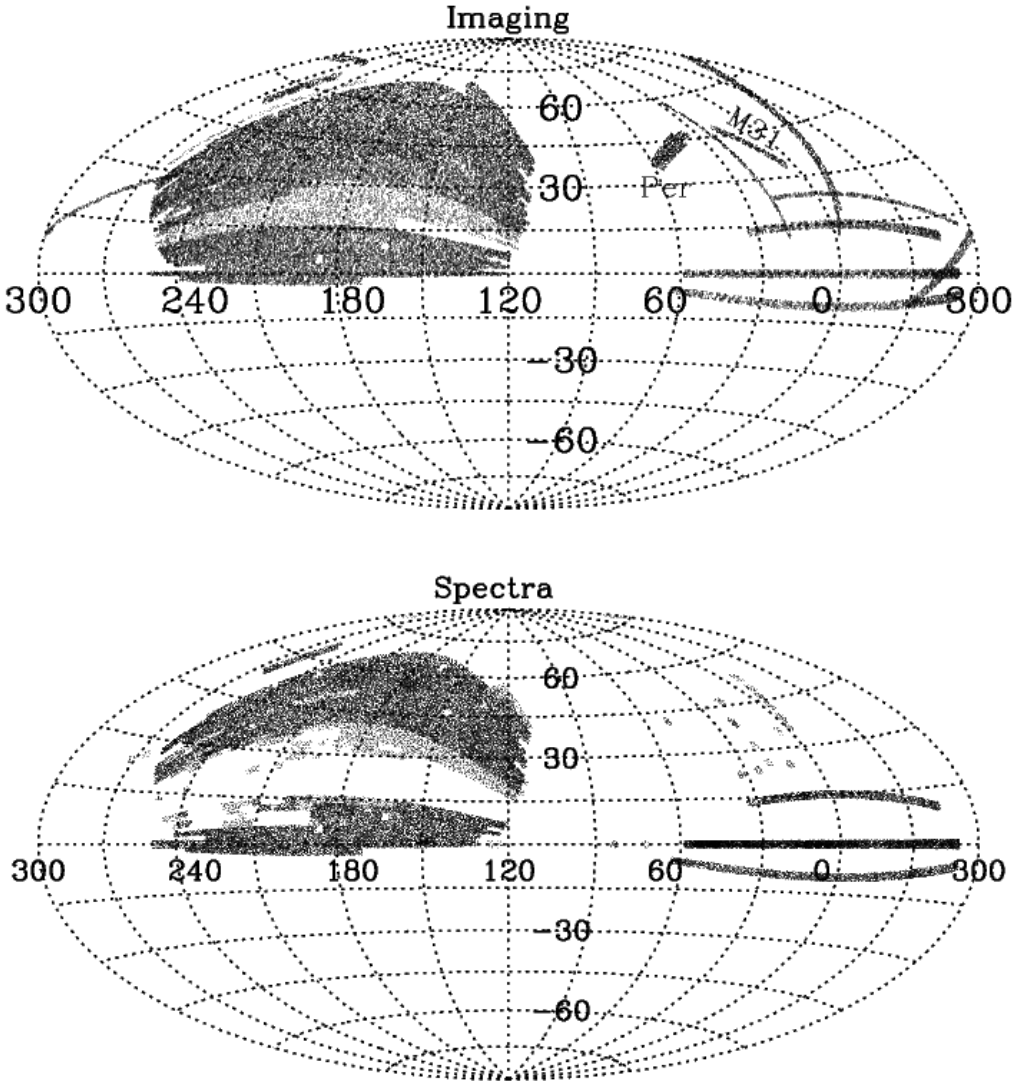


Figure 6.1: **Distribution on the sky of SDSS imaging**.
Distribution on the sky of SDSS imaging (*top*) and the spectroscopy (*bottom*) included in DR5, shown in $J2000.0$ equatorial coordinates.

Parameter	**Value**
Footprint area	$8000 deg^2$
Imaging catalog	217 million objects
Magnitude limits:	
u	$22.0 mag$
g	$22.2 mag$
r	$22.2 mag$
i	$21.3 mag$
z	$20.5 mag$
Median PSF width	$1.4''$ in r
rms photometric calibration errors:	
r	2%
$u-g$	3%
$g-r$	2%
$r-i$	2%
$i-z$	3%
Astrometry errors	$< 0.1''$ rms absolute per coordinate
Object counts:	
Stars primary	$85,383,971$
Stars, secondary	$28,201,858$
Galaxies, primary	$131,721,365$
Galaxies, secondary	$33,044,047$

Table 6.1: **Characteristics of DR5 Imaging Survey.**

D'Abrusco et al. (2007) with a neural network approach based on *Multi-Layer Perceptron*, or *MLP*. Neural Networks have long been known to be excellent tools for interpolating data and for extracting patterns and trends. In the last few years they have also dug their way into the astronomical community. In practice, a neural network is a tool which takes a set of input values (*input neurons*), applies a non-linear transformation and returns an output. The optimization of the output is performed by using a set of examples for which the output is known *a prior* [5].

6.2.1 Photometric and Spectroscopic Redshift Catalogue

In order to generate this catalogue two steps are adopted [5]: firstly, an MLP was trained to recognize nearby ($z < 0.25$) and distant ($z > 0.25$) ob-

Parameter	**Value**
Footprint area	$5713 deg^2$
Wavelength coverage	$3800 - 9200$ Å
Resolution $\lambda/\Delta\lambda$	$1800 - 2100$
S/N	$> 4 pixel^{-1}$ at $g = 20.2$
Redshift accuracy	$< 5 kms^{-1}$
Number of plates	$30 kms^{-1}$ rms for main galaxies
Number of spectra	$1,048,960$
Galaxies	$674,741$
Science Primary galaxies	$561,530$
Stars	$212,781$

Table 6.2: **Characteristics of DR5 Spectroscopic Survey.**
This table summarizes the characteristics of DR5 Spectroscopic Survey. The redshift accuracy refers to main galaxies, $\sim 99\%$ of classifications and redshifts are reliable. The science primary objects define the set of unique science spectra of objects from main survey plates (i.e., they exclude repeat observations, sky fibers, spectrophotometric standards.)

jects, than two separated MLPs were trained to work in these two different redshift regimes. Such approach finds a strong support in the fact that in the SDSS DR5 catalogue, the distribution of galaxies inside the two different redshift intervals is dominated by two different galaxy populations: the Main Galaxies (hereby, MG) sample in the nearby region, and the Luminous Red Galaxies (LRG) in the distant one. The use of two separate networks, ensures that the NNs achieve a good generalization capability in the nearby sample, leaving the biases in the distant. To perform the separation between MG and LRG objects, a spectroscopic sub-sample was extracted from SDSS-DR4.
After this step, the evaluation of photometric redshift was performed working separately, in the two regimes. The results are summarized as [6]:

- For the MG sample, the robust variance turned out to be $3\sigma = 0,028$ over the whole redshift range, $3\sigma = 0.0197$ and $3\sigma = 0.0245$ for the nearby and distant objects, respectively.

- For LRG sample the variance obtained is $3\sigma \approx 0.0163$ over the whole range, $3\sigma \approx 0.0154$ and $3\sigma \approx 0.0189$ for the nearby and distant objects, respectively.

6.3 Catalogue of Berlind

For evaluating the New FOF algorithm, its performance are compared against performance obtained by the FOF algorithm described by Berlind [3]. Firstly, it is worth to recall that the Berlind FOF algorithm is optimized to find galaxy groups and clusters in volume-limited samples of the Sloan Digital Sky Survey (SDSS), and in particular from a part of the Data Release 3. The authors of this work are interested in using galaxy groups to constrain the properties of galaxies as a function of their underlying dark matter halo mass. It is therefore important that the population of galaxies constituting the groups is homogeneous within the sample volume. Thus, this is the main reason why they chose to construct volume-limited subsamples of the full SDSS redshift sample that are each complete in a specified redshift range down to a limiting r-band absolute magnitude threshold.
To the extent of this thesis it is in fact more important to understand how important parameters such as the linking lengths are chosen. As mentioned, the most important ingredient of our group-finding algorithm is our choice for the linking lengths $b_{\parallel}$ and $b_{\perp}$. Since that the final aim of both, this thesis and Berlind's work, is to obtain a balance between being inclusive (groups together every galaxy inhabiting the same halo), and reducing contamination, while producing groups that have an unbiased multiplicity function, it is very interesting to compare the New FOF with the Berlind one, particularly about the linking lengths. Berlind finds for the linking lengths the following optimum values of $b_{\perp} = 0.14$ and $b_{\parallel} = 0.75$. Such values beside the mentioned analytical relationships, are refined making an extensive use of the mock catalogue, in order to achieve a catalogue composed by galaxy groups with N ¿= 10 that have

- an unbiased multiplicity function;
- an unbiased median relation between the multiplicities of groups and their associated halos;
- a spurious group fraction of less than $\sim 1\%$;
- a halo completeness (fraction of halos that are associated one-to-one with groups) of more than $\sim 97\%$;

- a velocity dispersion distribution that is $\sim 20\%$ too low at all multiplicities.

- the correct projected size distribution as a function of multiplicity;

6.4 The original FOF:running on the real data set

6.4.1 The Data

The evaluation started by letting the original FOF algorithm run on the spectroscopic subsample SDSS-DR4 which covers 6670 deg^2 on the sky and contains $\sim$ 449271 objects. Since the apparent magnitude limit of the redshift sample varied across the sky, in the commissioning phase of the survey, the r-band magnitude limit was cut to 17.5.
Then three volume-limited subsamples of the full DR4 spectroscopic data set are isolate, each of those complete in a specified redshift range down a limiting r-band absolute magnitude. Hence, the three samples are built by selecting galaxies within three redshift range: $0.015 < z < 0.1$, $0.015 < z < 0.068$,$0.015 < z < 0.045$, within limiting absolute magnitudes equal to M_{r20}, M_{r19}, M_{r18} respectively. The absolute magnitudes are quoted for $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $h \equiv H_0/100 kms^{-1}Mpc^{-1} = 0.7$.

In the figure 6.2, all galaxies belonging to the three samples are shown in the luminosity redshift plane.

The sharp cutoff curve along the lower right part of the plot, shows the survey apparent magnitude limit $r = 17.5$. The limiting absolute magnitude of each sample changes slightly with z due to evolution corrections applied to galaxies luminosity: as a galaxies moves to the outer edge of a given volume-limited sample, its luminosity somewhat increases, allowing lower redshift galaxies to make it into the sample at lower luminosities that they do at higher redshift. The maximum redshift value was chosen since this yields the largest possible volume limited samples; the value of z_{min}, used for all three samples, is chosen to alleviate some of the problems associated with obtaining accurate photometry of nearby extended galaxies.

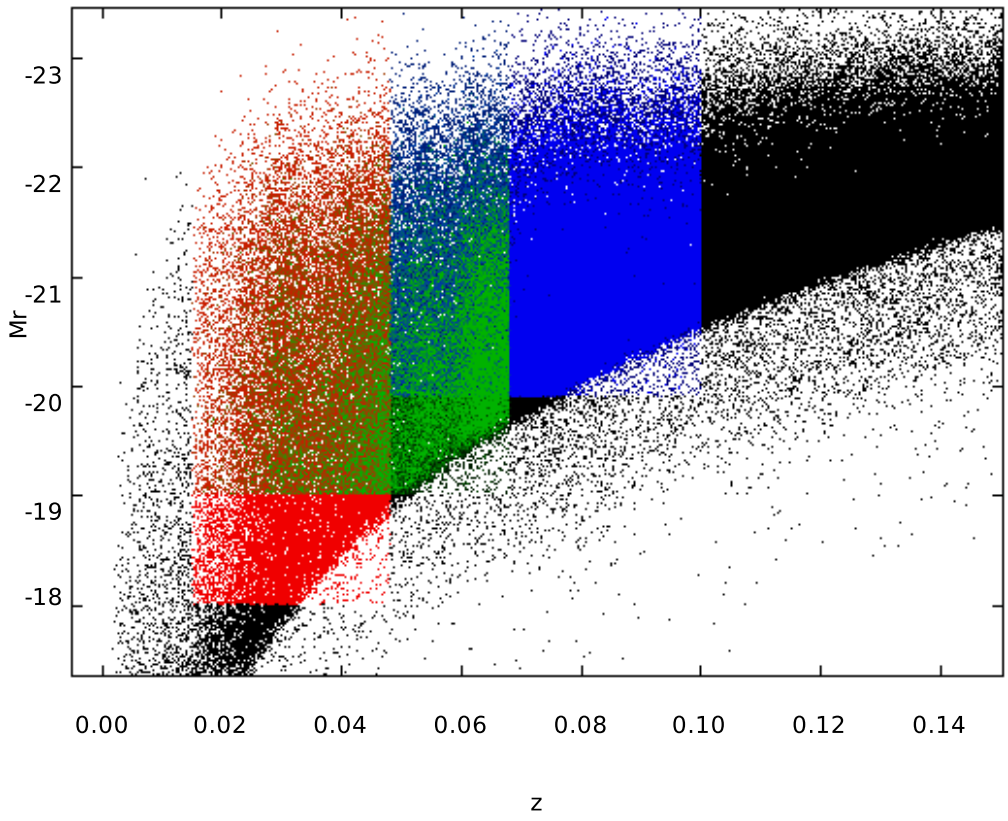


Figure 6.2: **Absolute r-band magnitude vs. redshift for galaxies in SDSS redshift survey**.
M_{r20} (*blue*), M_{r19} (*green*), M_{r18} (*red*) show the three catalogs and each dot show the three spectroscopic subcatalogues used in this word.

Name	z_{min}	z_{max}	$< M_r$	N_g	$\bar{n}_g$
M_r18	0.015	0.048	18	39253	$0,0400$
M_r19	0.015	0.068	19	72722	$0,0213$
M_r20	0.015	0.100	20	140233	$0,0131$

Table 6.3: **Volume-limited Sample Parameters.**
This table summarizes the parameters of the volume limited samples. In the col.2 and 3 are the redshift limits of each sample. In col.4 is the limiting magnitude. In the col.5 and 6 the number of galaxies and the galaxy density in units of $h^3 Mpc^{-3}$ for each sample.

6.4.2 Group and Cluster Spectroscopic Catalogue

The original friends of friends algorithm was applied on each of the three volume limited samples.
According to the results obtained on mock simulations, the linking lengths are chosen to be equal:

$$D_L \leq b_\perp \bar{n}^{-1/3} = 0.2\bar{n}^{-1/3}$$

in the transverse direction, and

$$V_L \leq b_\parallel \bar{n}^{-1/3} = 0.5\bar{n}^{-1/3}$$

in the line of sight directions. The resulting group catalogues for the three samples contain a total of 8116, 4760 and 2420 groups,respectively. For each group, an unweighted centroid is calculated, as the average right ascension, declination and mean redshift of all galaxies belonging to the group. An rms projected radius is also estimated as:

$$R_{\perp,rms} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} {r_i}^2} \tag{6.1}$$

The results are shown in table 6.4. In the Figures 6.3,6.4,6.5, the multiplicity function of Mr18 sample is shown.

Name	N_{gr}	N_{gal}	$Richness_{Max}$
M_r18	2420	13789	470
M_r19	4706	26125	312
M_r20	8116	39922	184

Table 6.4: **Results obtained.**
In the col. 2 there are the number of groups obtained, in the col. 3 it is shown the number of galaxies within groups and in the col4. are listed the maximum richness of catalogues.

6.5 The New FOF applied to a photometric data set

We used the photometric sample by D'Abrusco et al. (2007) From that, we extracted three samples M_{r18}, M_{r19}, M_{r20} with the same selection criteria

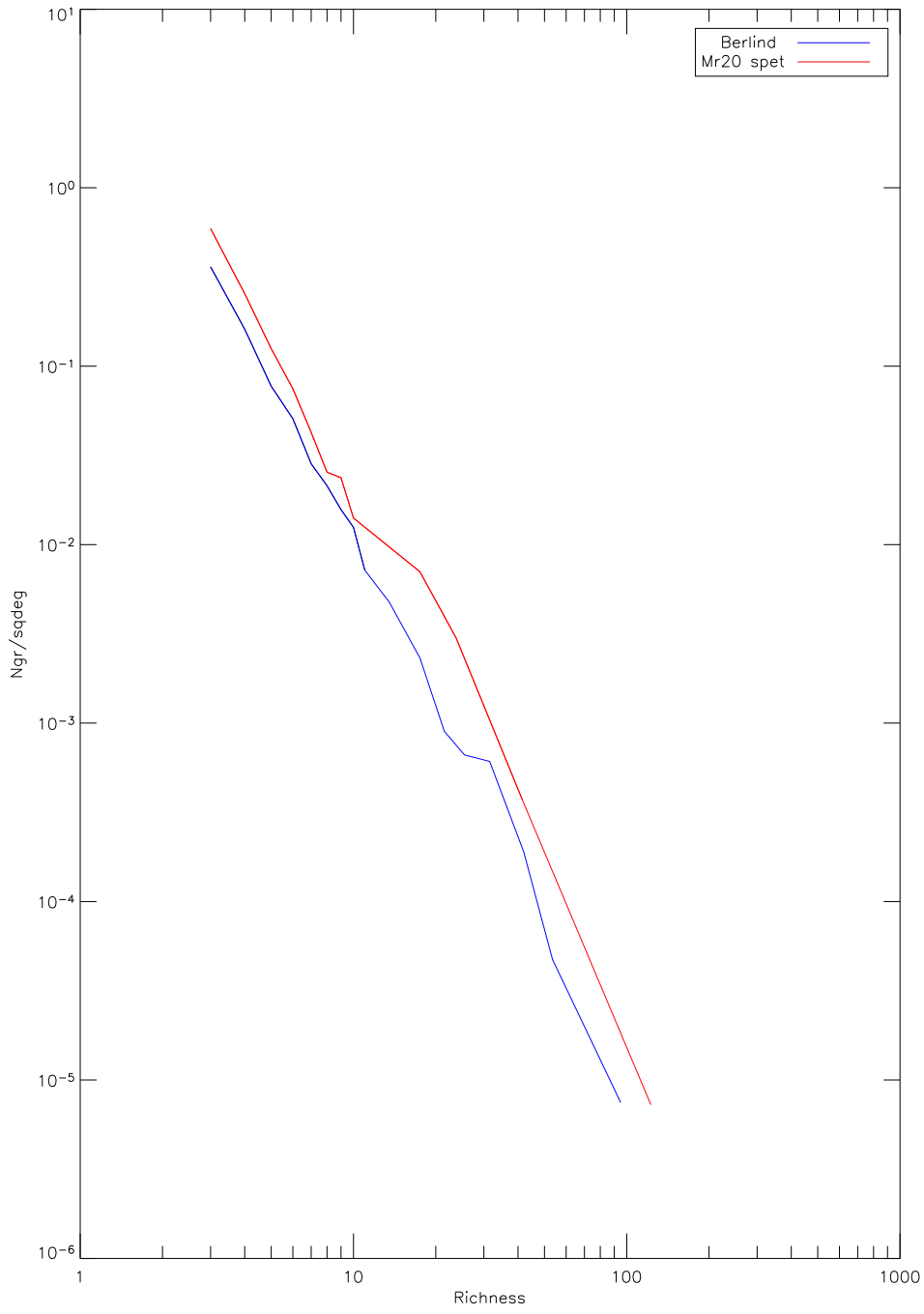


Figure 6.3: **Multiplicty Function Mr20 sample**.
The multiplicity function of Berlind's catalogue [3] is shown too.

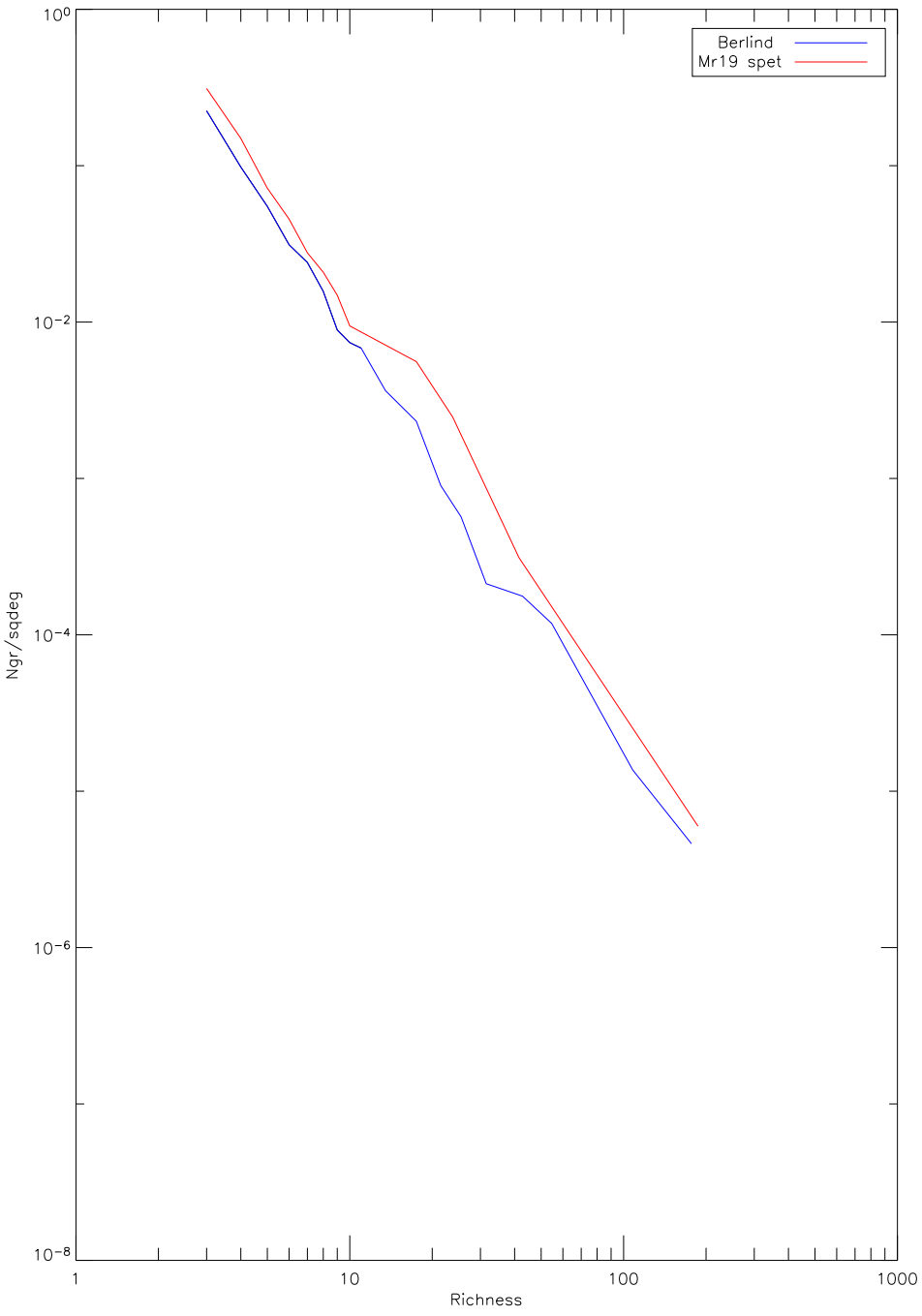


Figure 6.4: **Multiplicty Function Mr19 sample**.
The multiplicity function of Berlind's catalogue [3] is shown too.

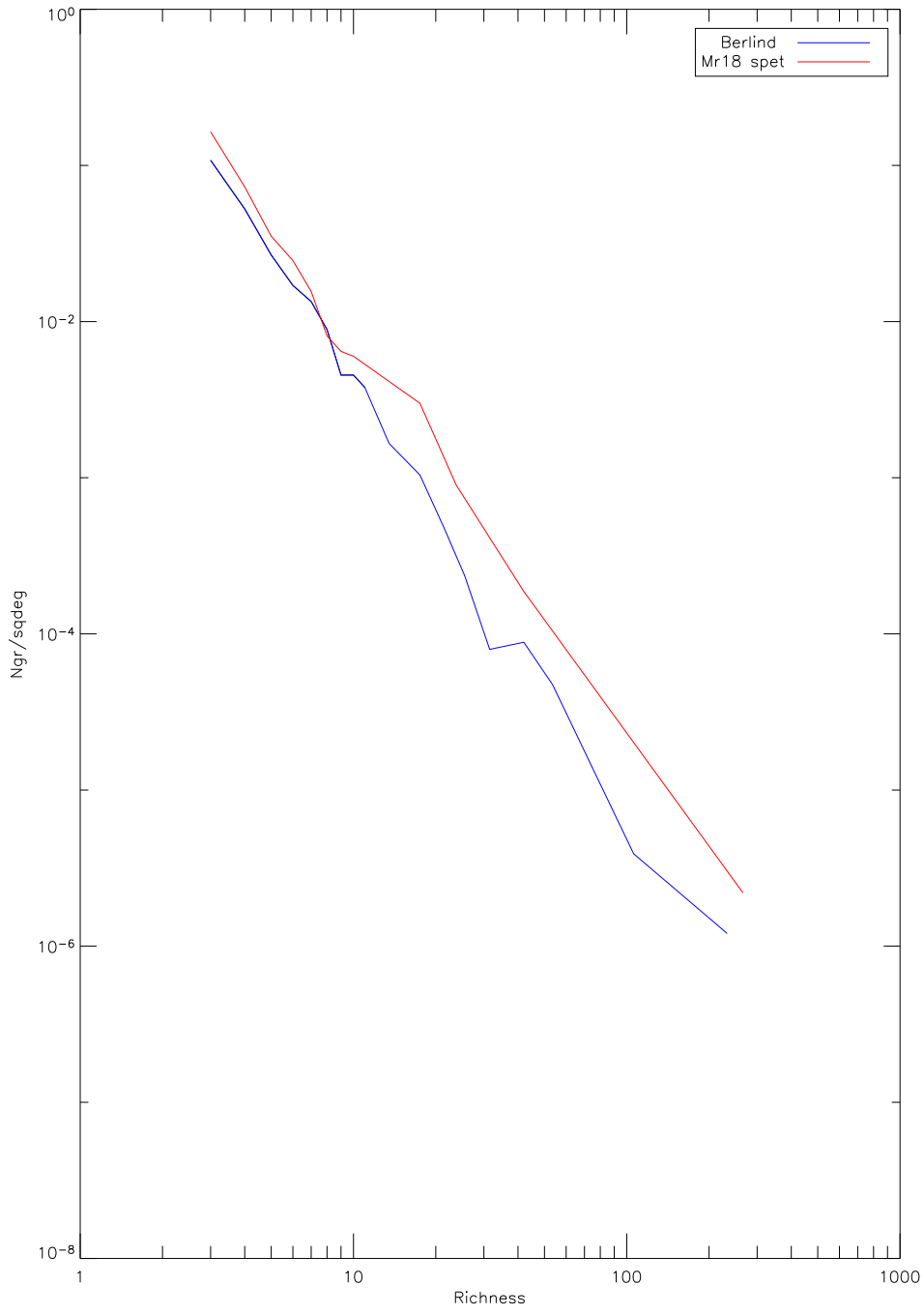


Figure 6.5: **Multiplicty Function Mr18 sample.**
The multiplicity function of Berlind's catalogue [3] is shown too.

as for the spectroscopic data set. The absolute magnitudes of each galaxies were calculated using the same cosmology above, but were calculated using

$$m_r - M_r = 5 \log D_z - 5 \tag{6.2}$$

where m_r and M_r are respectively the apparent and absolute magnitude, and d is the object distance as $D_z = H_0^{-1} c z_{phot}$. The linking lengths are chosen to be equal to:

$$D_L \leq b_\perp \bar{n}^{-1/2} = 0.16 \bar{n}^{-1/2}$$

in the transverse direction, and

$$V_L \leq b_\parallel \bar{n}^{-1/2} = 1 \bar{n}^{-1/2}$$

in the line of sight directions (see, Cap. 5). It is important to recall that when dealing with photometric redshifts, photometric errors are taken into account, so the linking condition along line of sight becomes:

$$V_i \leq [(V_L/2^2 + (v\delta z_i)^2]^{1/2} \leq \frac{V_L}{2} \tag{6.3}$$

$$V_j \leq [(V_L/2^2 + (v\delta z_j)^2]^{1/2} \leq \frac{V_L}{2} \tag{6.4}$$

where $\delta z_i = 0.02$. Making use of the values calculated from the simulated data such as $b_\perp = 0.4$, the finding algorithm generates very long galaxies chains that result not physically bound objects. Thus, in attempt to solve the mentioned problem, the trend of the projected linking length was been evaluated considering $b_\perp = 1$. From the application of this methodology, it is been observed that the linking length value increases for decreasing slice density. This effect is due to the well known fact that while the redshift value increases, the density of the particles decreases, hence the average distance between objects within the group grows up in value. Making use of above mentioned value for the parameter $b_\perp = 1$, it is been observed that the linking length reaches the value of 7 Mpc. Therefore, several values for the linking length (starting from 1 down to 0.1) were been evaluated before the proper value was found: $b_\perp = 0.16$. In the Fig 6.6 the trend of this linking length is shown as function of z-slice. The results are shown in the Table 6.5.

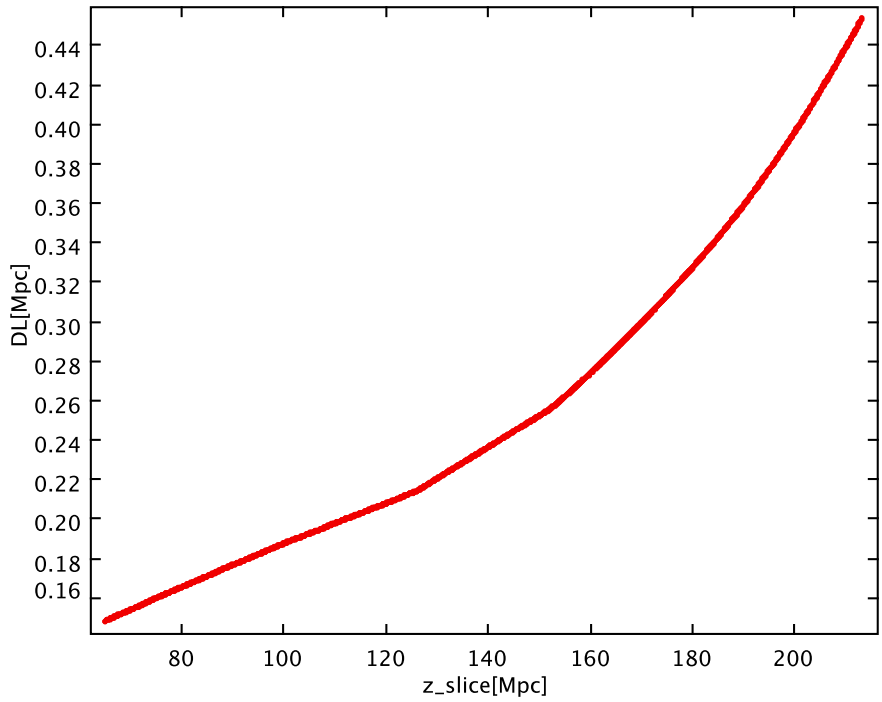
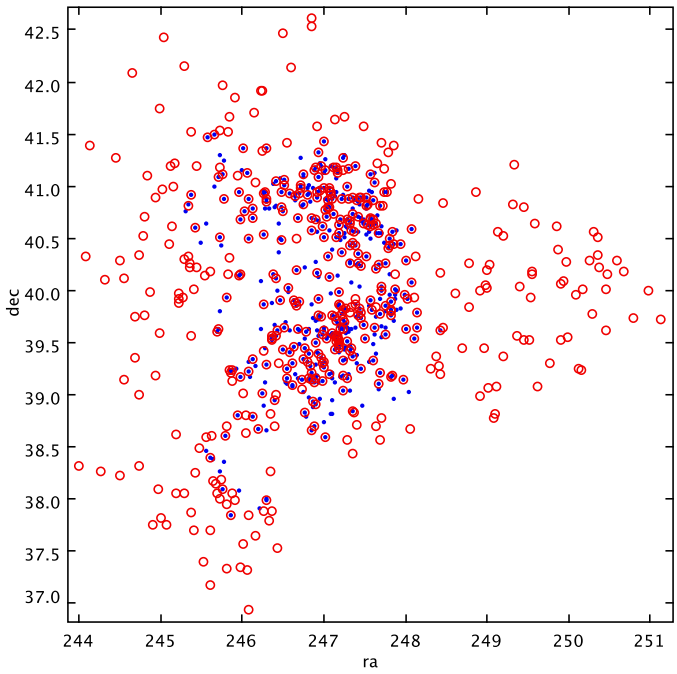


Figure 6.6: **Trend of D_L as function of z slice.**

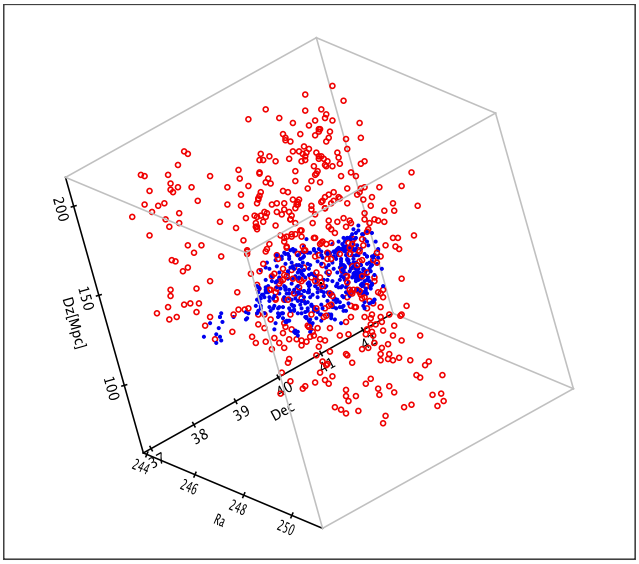
σ	N_{gr}	N_{gal}	$\mathcal{R}_{max}$	**K-S**	
				D	**P**
0.3	1347	9266	610	0.09	96%
0.4	1280	9027	487	0.14	67%
0.8	1338	9890	517	0.18	32%
1	1338	9890	517	0.18	32%
3	1595	9767	395	0.11	93%

Table 6.5: **The photometric groups results.**
This table summarizes the results obtained by New FOF on the M_r18 volume limited sample with $b_\perp = 0.16$. These results were obtained at different σ values. In the column 2 and 3 there are the number of groups and the number of galaxies within. In the column 4 is the maximum value of richness in the catalogue. For each catalogue, Kolmogorov-Smirnov statistical test was performed, in order to compare the both MF of M_r18 photometric catalogue with MF of the M_r18 spectroscopic one. The results are shown in the column 3. In the "D" column, the maximum deviation between comulative distribution of the data and supplied function is shown. In the "P"column, there are hte probabilities that the two distribution of data are drawn from the same distribution.

In figure 6.7, the comparison of the richest groups in the two catalogues (the photometric catalogues (*red*) and the spectroscopic one (*blue*). The multiplicity Function of each catalogue was proved (see fig 6.8 and 6.9). Whereas the spectroscopic data set is a subsample of the DR4, while photometric one is derived from DR5, the two multiplicity functions have different standards. However, from a comparison of their trend it is possible to observe that NEW FOF reproduces results in agreement with those derived from original FOF.

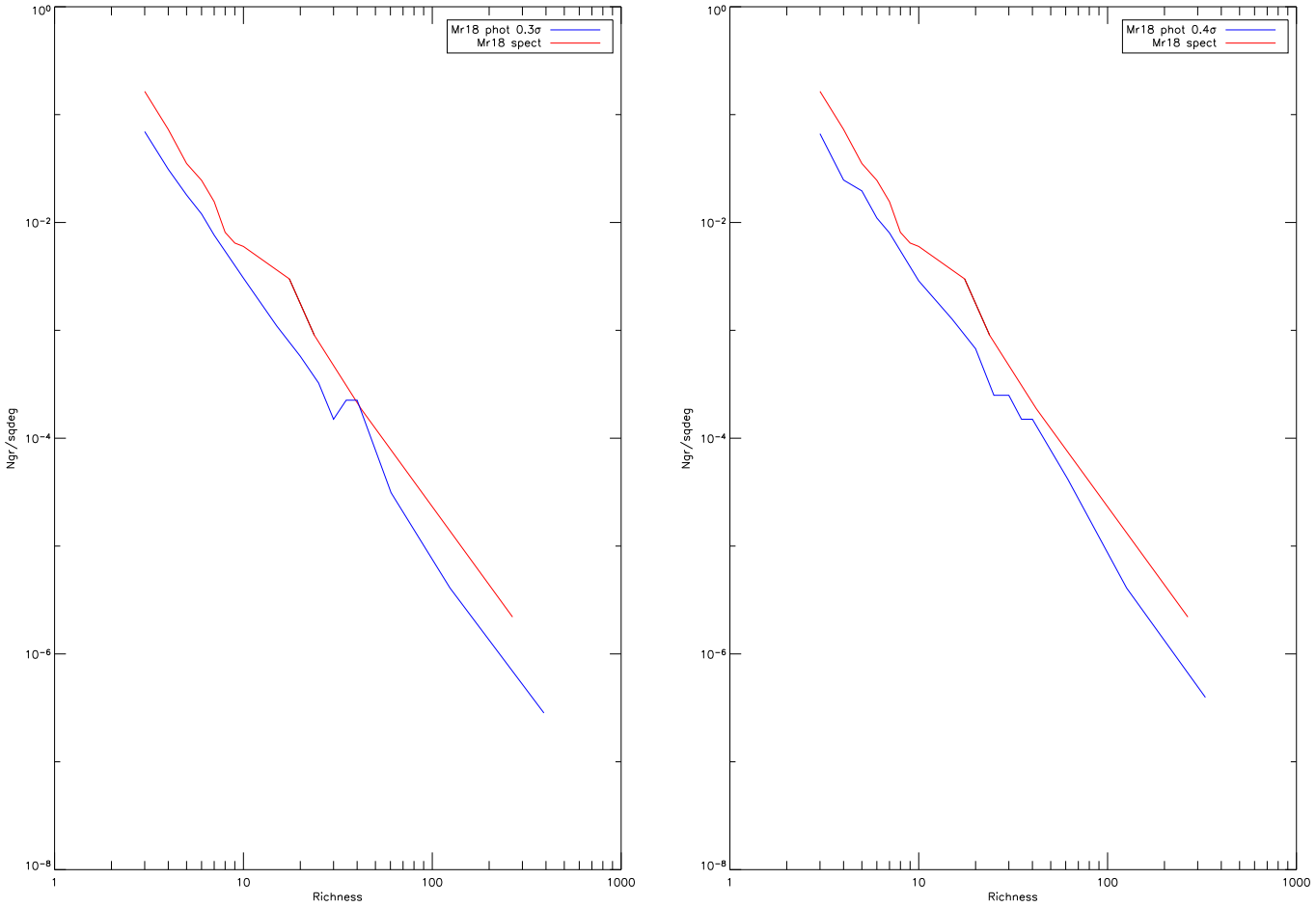


(a)



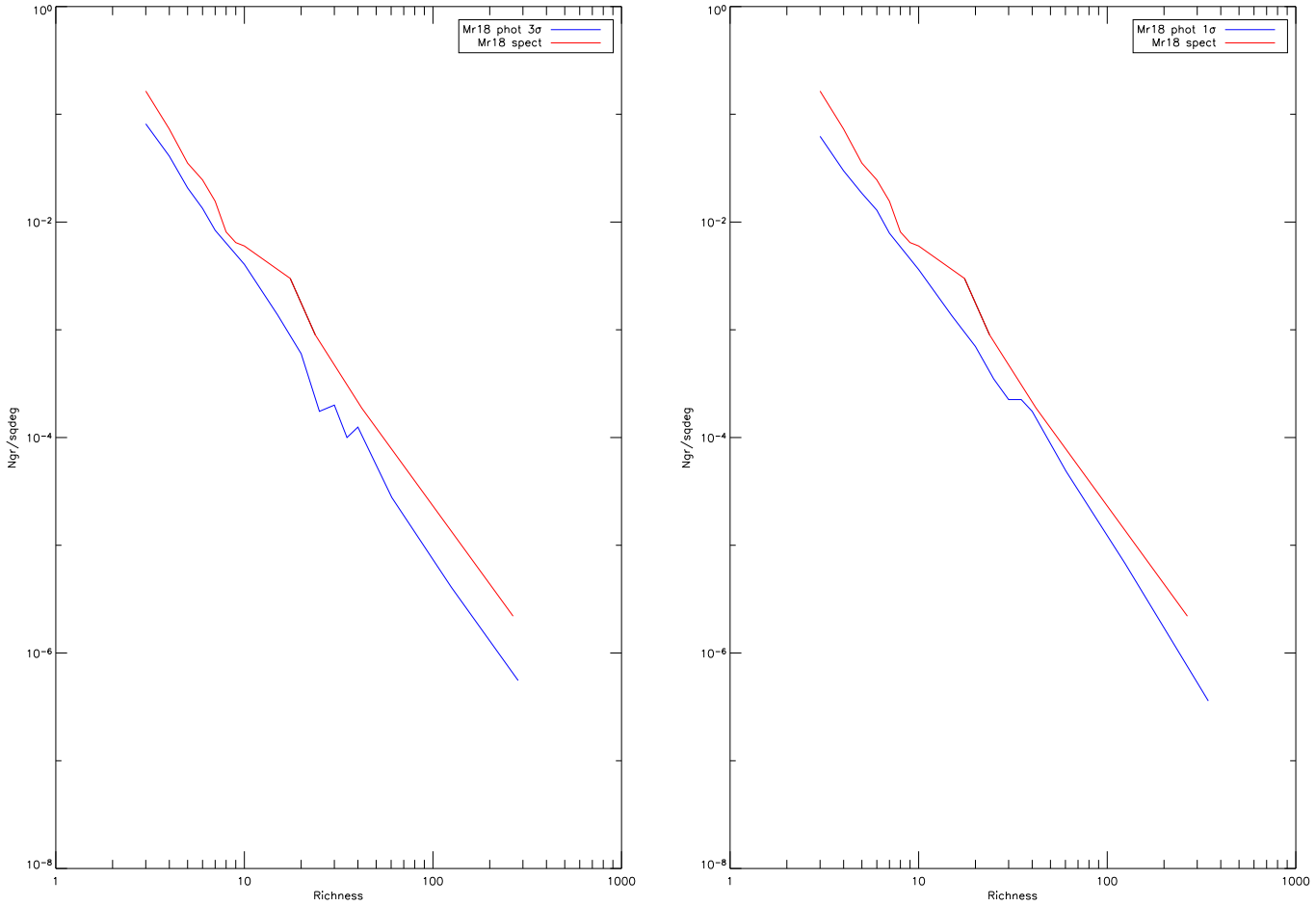
(b)

Figure 6.7: **Comparison between the richest groups in the two catalogues.**
The figure a) shows the projection of galaxies in the sky. The red circle represents the group ID35 of the $M_r18$photometric catalogue and the blue points are the ra and dec coordinates of the ID1802 of the $M_r18$spectroscopic catalogue. The figure b) shows a plot in 3D of the same groups: the red circles are the ra, dec and Dz of the ID35 photometric group and the blue points are the same coordinates of ID1802 spectroscopic group.



(a) (b)

Figure 6.8: **Multiplicity Function Mr18 photometric sample**.
The groups catalogue was generated by original FOF with $b_{\parallel} = 1$ and $b_{\perp} = 0.16$. In the fig a) is shown the MF of catalogue obtained with 0.3σ and in fig. b) the MF at 0.4σ.



(a) (b)

Figure 6.9: **Multiplicity Function Mr18 photometric sample**.
The groups catalogue was generated by original FOF with $b_{\parallel} = 1$ and $b_{\perp} = 0.16$. In the fig a) is shown the MF of catalogue obtained with 3σ and in fig. b) the MF at 1σ.

Summary and Conclusions

In this thesis, a new finding algorithm for galaxies structures and clusters was been described, together with the complete set of necessary tests, starting from simulated data, up to an application to real ones.
The choice for the name of the presented algorithm (New FOF) it is not casual, this is in fact an evolution of the well known FOF (Friend Of Friends) finding algorithm, presented for the first time by Huchra & Geller in 1982s [10]
The performances of the algorithm and of course, the necessary choice of parameters, as recommended by the literature, was performed as follow: firstly, the original FOF algorithm is applied to a spectroscopic mock data set in order to determinate the values of the most important set of parameters: the linking lengths. Then, the application of the NEW FOF to the same sample data set, allows the selection of the $z-$slicing step and the number of slices to be created within the data sample.
However, as it is possible to observe from the discussion held in Chapter 5, the passage from spectroscopic data photometric ones, has some huge implications that must be considered in the parameters choice. It is in fact worth to highlight that it is not possible to use only one pair of linking lengths for both algorithms. Namely, because in the original FOF, the linking length remain constant (there is in fact no dependence with redshift), for the NEW FOF there is a strong dependence of it with z.
Recalling that for the original FOF the projected linking length is defined as

$$D_L = b_\perp \bar{n}_g^{1/3} \qquad V_L = b_\parallel \bar{n}_g{}^{1/3}$$

where $\bar{n_g}$ is the mean number density of galaxies in the sample while, for the NEW FOF linking length becomes

$$D_L = b_\perp \bar{n}_{gs}^{1/2}$$

where $\bar{n}_{gs}$ the surface density of each $z-$slice and the linking condition on the line of sight becomes

$$V_i \leq [(V_L/2^2 + (v\delta z_i)^2]^{1/2} \leq \frac{V_L}{2}$$

where δz_i is the individual redshift errors equal to 0.02 and V_L is defined

$$V_L = b_\parallel \bar{n}_{gs}^{1/2}$$

Data evaluation on both simulated and real data sets needs to take into account the mentioned observation. The resulted spectroscopic catalogue obtained running the Original FOF on the mock catalogue of a completeness of $\sim$ 99%, while the catalogue obtained running with New FOF on the photometric mock data set is complete to $\sim$ 80%.

Logically, the next step of the algorithm evaluation was run it on real data set. In order to compare the results with those obtained from spectroscopic data, the same cuts in magnitude and redshift were applied. For each catalogue of spectroscopic structures, the Multiplicity Function is calculated and compared with Berlind's groups catalogues. Although there are differences between the catalogs derived in this work and those in the literature, these are due to different choices of Linking Lengths, but primarily to the fact that the samples studied are different. Berlind searched for groups on NYGC that is a subsample of SDSS-DR3, while in this work the real data are derived from a subsample of DR4 (spectroscopic redshifts). The photometric redshift catalogue used for this work, has been constructed by D'Abrusco et al. (2007) with a neural network approach based on *Multi-Layer Perceptron*, or *MLP*. The results of M_{r18} photometric subsamples, obtained by running NEW FOF, were compared with the M_{r18} spectroscopic one. In conclusion, given the results obtained from the trend evaluation of both the Multiplicity Functions as shown in chapter 6 it is possible to say that the presented algorithm clearly fulfill the targeted aim of extending the capabilities of the

original FOF algorithm to photometric redshifts datasets without any loss in the mentioned (Chapter 3), important peculiarity and features that made of the original FOF one of the election methodologies in the structure finding branch of modern Cosmology.

Bibliography

[1] Agekyan, T. A., Anosova, Zh. P., 1967, SvA,44,1261

[2] Bahcall, N.A., 1996 astro-ph/9612046

[3] Berlind, A. A., Frieman, J., Weinberg, D. H., et al. (the SDSS Collaboration), 2006, ApJS, 167, 1

[4] Botzler,C.S., Singula, J., Bender,R., Hopp, U., 2004,MNRAS,349..425B

[5] D'Abrusco, R., Staiano, A., Longo, G., Brescia, M., Paolillo, M., De Filippis, E., Tagliaferri, R., 2007,Ap.J.,663-752

[6] D'Abrusco, R., Staiano, A., Longo, G., Brescia, M., Paolillo, M., De Filippis, E., Tagliaferri, R., 2007,astro-ph/0701135v2

[7] Gott,J.R., Turner,E. L, 1977,Ap.J.,212:357-371

[8] Governato, F., Tozzi, P., Cavaliere, A., 1996. Ap.J. 458,18

[9] Hickson, P.,1982, Ap.J.,255,382

[10] Huchra, J. P. & Geller, M.J., 1982, Ap.J.,257:423-437

[11] Liddle, Andrew, *An Introduction to Modern Cosmology*, *University of Sussex,UK*

[12] Mamon,G.A., 1996,Ap.J.,307,426

[13] Press, W. H., Schechter, P., 1974, Ap.J,187:425-438

[14] Sanchez, F., Collados, M.,Rebolo,R., *Observational and Cosmology*, Cambridge

[15] Turner,E. L,Gott,J.R., 1976,Ap.J. Suppl,32:409-427

[16] Zheng, J.Q., Valtonen, M. J., Chernin, A. D., 1993, AJ,105,2047Z